

1956

# Components of variance of diallel crosses of maize in experiments repeated over locations and years

Dale Frederick Matzinger  
*Iowa State College*

Follow this and additional works at: <https://lib.dr.iastate.edu/rtd>



Part of the [Agriculture Commons](#), and the [Plant Pathology Commons](#)

---

## Recommended Citation

Matzinger, Dale Frederick, "Components of variance of diallel crosses of maize in experiments repeated over locations and years" (1956). *Retrospective Theses and Dissertations*. 14727.  
<https://lib.dr.iastate.edu/rtd/14727>

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact [digirep@iastate.edu](mailto:digirep@iastate.edu).

COMPONENTS OF VARIANCE OF DIALLEL CROSSES OF MAIZE IN  
EXPERIMENTS REPEATED OVER LOCATIONS AND YEARS

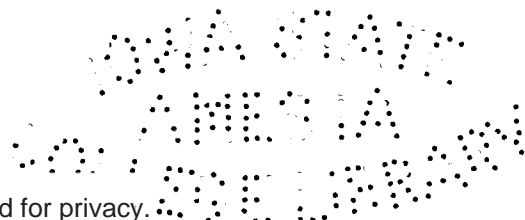
by

Dale Frederick Matzinger

A Dissertation Submitted to the  
Graduate Faculty in Partial Fulfillment of  
The Requirements for the Degree of  
DOCTOR OF PHILOSOPHY

Major Subject: Crop Breeding

Approved:



Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

Professor in Charge, Farm Crops

Signature was redacted for privacy.

Head of Major Department

Signature was redacted for privacy.

Dean of Graduate College

Iowa State College

1956

UMI Number: DP14596

### INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

**UMI<sup>®</sup>**

---

UMI Microform DP14596

Copyright 2006 by ProQuest Information and Learning Company.

All rights reserved. This microform edition is protected against unauthorized copying under Title 17, United States Code.

ProQuest Information and Learning Company  
300 North Zeeb Road  
P.O. Box 1346  
Ann Arbor, MI 48106-1346

SE 131, 28

M 4332

## TABLE OF CONTENTS

	<u>Page</u>
INTRODUCTION . . . . .	1
REVIEW OF LITERATURE . . . . .	3
THEORETICAL DERIVATION OF COVARIANCES . . . . .	17
CONDUCT OF EXPERIMENTS . . . . .	66
ANALYSIS OF EXPERIMENTS . . . . .	69
RESULTS . . . . .	89
DISCUSSION . . . . .	96
CONCLUSIONS AND SUMMARY . . . . .	105
LITERATURE CITED . . . . .	108
ACKNOWLEDGEMENT . . . . .	113

772169

## INTRODUCTION

In a practical corn breeding program the isolation of inbred lines is directed towards obtaining those lines which will contribute favorable genes or combinations of genes to their resulting hybrid combinations. The method of isolation of these lines and the combination of them into hybrids is based upon fundamental laws of genetics.

Since many of the desired characters are controlled by genes at more than a single locus, quantitative characters, one is not able to measure the direct effect of each gene substitution. For this reason one must use statistical procedures to obtain basic genetic information. Estimates of certain genetic parameters will aid an investigator in choosing the most efficient breeding method.

A set of diallel crosses is the  $n(n-1)$  possible combinations from  $n$  lines, such that every line is crossed with each of the remaining  $n-1$  lines. In this thesis the general theory of diallel crosses is presented with methods given for the estimation of certain useful genetic parameters. The general case of arbitrary inbreeding, arbitrary number of alleles per locus, and arbitrary epistacy is presented for two loci with an indication of the extension to arbitrary loci.

Estimates of several of the useful genetic components are obtained from diallel crosses of a group of random  $S_1$  lines. The crosses were

repeated at three locations in each of two years to obtain information on the consistency of the estimates. Estimates of interactions of the genotypic effects with districts and years are given. An attempt is made to give biological meaning to the statistical estimates of the parameters in the original model.

## REVIEW OF LITERATURE

## Correlation between Relatives

Population genetics had its origin soon after 1900. Pearson (44) gave the first account of the correlation between relatives when he considered only two alleles and complete dominance. At that time he was not able to explain observational data on the basis of Mendelian segregation. Yule (65) showed that the observations examined by Pearson were explainable by assuming the absence of dominance.

A complete demonstration of the multiple factor theory of quantitative inheritance was given by Nilsson-Ehle (43) in 1908 on wheat and by East (11) in 1910 on corn. The study of the inheritance of qualitative characters on an individual cross basis had to be modified for the study of continuous variation to the analysis of frequencies and proportions of individuals from some base population.

In 1908, Hardy (19) and Weinberg (58) independently gave the basic law specifying the property of equilibrium for a random mating population. They showed that there is no change in the genotypic proportions of a random mating population from generation to generation and therefore no change in gene frequency.

Much of the early work assumed that genes combined in an additive manner. Weinberg (58,59) was the first to consider the subdivision of

phenotypic variability into a genotypic and an environmental contribution. An arbitrary number of alleles were considered. The method included dominance only for the case where the heterozygote was at some definite point between the two homozygotes. Some of the two-factor classical epistatic cases were considered.

Fisher (15) gave a complete treatment for any degree of dominance for random mating populations, giving correlations for parent-offspring, full-sib, uncle-nephew, cousin, and double-first-cousin relationships. Two-factor epistacy was considered for parent-offspring and full-sib correlations. The general solution included an arbitrary number of alleles. Special consideration was given to assortative mating and to the effects of linkage.

Wright (61) used the assumption of equilibrium in a random mating population and proposed the method of path coefficients for studying the correlations between relatives as a simplification to the system of determining population change on a gamete basis. Primarily the additive effects of genes were considered, so the covariances with respect to dominance deviations were not discovered by this method. Some consideration was given to dominance but it did not cover the general case. Assortative mating was considered for two alleles per locus having a substitution effect the same for all loci with no dominance.

A special type of gene action, the optimum number model, was proposed by Wright (62) in a correlation between relatives study. In considering a primary and secondary scale the grade on the primary scale



was determined by additive combinations of genes. On the secondary scale the optimum desirability was associated with an intermediate value on the primary scale. Wright (63) gave a review of his work on quantitative inheritance in 1951.

From a group of lines in all possible crosses Hull (26) obtained the partial regression of offspring on parent within a group having one parent in common as a relative measure of the heritability within the group. The assumption was made that epistacy was not important for yield of corn. The method was presumed to provide a measure of both the direction and magnitude of dominance. From 25 sets of corn data estimates of dominance were obtained over the entire range from dominance for low yield to overdominance. Comparisons were made between parents and  $F_1$ 's when the parents were not growing at the same location as the  $F_1$  generations.

The method of Mather (40) assumed no epistacy and a scaling method was proposed to satisfy this assumption. A scale was proposed to eliminate epistacy, and at the same time make the contribution of non-heritable agents independent of the genotype. No attempt was made to detect the type of gene action present. The amounts of additive (D), dominance (H), and environmental variance were estimated. The partition of variance considered the specific case of crossed and selfed populations of an  $F_1$  cross of two homozygous inbred lines where the gene frequency was one-half. The number of factors was estimated.

Mather and Vines (41) found that estimates of D and H for Nicotiana

rustica did not include all of the interactions observed for plant height. Evidence was given to show that genes for plant height interacted with each other and varied in their interactions in different environments. No method was available to measure the interactions, and interpretation of them was not possible.

In this type of experiment the assumption was made that estimates of environmental variance were the same for parents and  $F_1$  generations in comparison to  $F_2$  and backcross generations. Byrd (4) obtained estimates of within-plot variances for both parents and their  $F_1$  generations which varied within a given experiment and which interacted with locations and years. The data indicated that different scales would be needed for each cross and each environment. Gardner, et al. (16) indicated the importance of genotype-environmental interactions in corn. Interactions of this type result in an over-evaluation of the genetic improvement expected from selection in a single environment, as shown by Robinson, et al. (47).

Powers (45) found that in some tomato hybrids the environmental variability and the genotypic variability did not follow the same scale of measurement. In some of the crosses environmental variability was found to be arithmetic and in others it was logarithmic, with an analogous situation for the genotypic variability. In a later discussion, Powers (46) considered the effect of gene interactions and found that no one scale could be used to describe the genetic variability. A model was presented to obtain information on the main effects and interaction of

two pairs of genes.

New gene models originated which included terms for the epistatic parameters. Griffing (18) included an epistatic term which could not easily be extended to an arbitrary number of loci.

Anderson and Kempthorne (1) presented a factorial model, similar to that in the design of experiments, which included epistatic parameters for any number of loci. The method was outlined for two loci and extended to the general case of arbitrary loci. The model was applied to the means of populations obtained by successive selfing of a single individual and to populations arising from crossing two inbred lines followed by subsequent crossing and selfing. An application of the results to the scaling tests of Mather indicated that his tests are not exact tests of epistacy.

Hayman and Mather (22) extended the earlier work of Mather to include two-factor epistacy for the special case of crosses between inbred lines,  $p = \frac{1}{2}$ , and for only two alleles per locus.

Malécot (39) considered the case of random mating for a single locus with an arbitrary number of alleles and arbitrary dominance. The covariance between two individuals was shown to be

$$\frac{\phi + \phi'}{2} \sigma_A^2 + \phi \phi' \sigma_D^2$$

where  $\sigma_A^2$  is the additive genetic variance and  $\sigma_D^2$  is the dominance variance. The quantities  $\phi$  and  $\phi'$  characterize the relationship between the two individuals.

Kempthorne (34,35) has extended the results of Malécot to the general case of a random mating population with an arbitrary number of loci, arbitrary number of alleles, arbitrary dominance, and arbitrary epistacy with the assumption of no linkage effects. The general result obtained is that the coefficient of  $\sigma_X^2$ , where X contains r A-terms and s D-terms in the covariance of the two relatives is

$$\left[ \frac{\phi + \phi'}{2} \right]^r \cdot \left[ \phi \ \phi' \right]^s .$$

Kempthorne (36) presented correlations between relatives for the case of one locus in a population undergoing inbreeding by full-sibbing. Covariances between full-sibs and parent-offspring were shown to depend upon latent roots of the generation matrix. The general approach considered two alleles per locus with an indication of the extension to the case of an arbitrary number of alleles.

Horner and Kempthorne (24) determined the covariance between relatives for symmetrical random mating populations with gene frequency of one-half. Formulae for components of genotypic variance were given for the complementary, duplicate-factor, multiplicative, and optimum number models with a discussion of the consequences of each.

Cockerham (5) presented a method to obtain estimates of correlations between relatives with respect to epistatic contributions for random mating populations with an arbitrary number of loci, but with only two alleles per locus. Special consideration was given to the covariance

between relatives for crosses of inbred lines obtained at random from a random mating population. An exact solution was given for the case where the loci have recombination values of one-half. The results were given for full-sibs and half-sibs for arbitrary number of loci and arbitrary epistacy, but considered only two alleles per locus.

In all of the studies reported in the literature, there has been no general treatment of the effect of linkages on the estimation of correlations between relatives. Most of the methods assume that the population is in linkage equilibrium. Cockerham (6) showed that covariances between relatives are affected by linkage when one individual is not a common ancestor of the other, even though the genotypic frequencies of the population are in linkage equilibrium. The amount of bias is expressed only in the epistatic components of the model, and increases in magnitude as the recombination frequency at a locus decreases.

#### Degree of Dominance

The term heterosis as proposed in 1914 by Shull (54) was intended to avoid any implication as to the exact mechanism involved in causing the increased vigor observed from inbreeding and crossing maize. Through common usage, heterosis and hybrid vigor have been used to describe the increase in performance of an  $F_1$  over the best performing parent.

Two quite different types of gene action have been postulated to account for heterosis. These are:

1. Dominant, favorable growth factor hypothesis

## 2. Divergent alleles or overdominance hypothesis.

Davenport (9) indicated that dominant genes were more often beneficial to an organism than recessive genes. Bruce (3) gave a mathematical approach showing algebraically that a hybrid population would always contain more dominant phenotypes than any subsequent random mating population. Keeble and Pellew (31) showed that hybrids between two pure varieties of peas combined long internodes from one parent and a larger number of internodes from the other parent to give a hybrid taller than either parent. In 1917 Jones (30) added the assumption of linkage to the general hypothesis of Bruce. He showed that if a favorable dominant gene and a detrimental recessive were closely linked, the heterozygous chromosome would be superior to both homozygotes.

Shull (52) gave the first reference to the physiologic stimulation hypothesis by indicating an increase in stimulation, measured by increased vigor, with an increase in the diversity of uniting gametes. Shull (53) and East and Hayes (13) proposed the "physiological stimulation hypothesis", attributed hybrid vigor to heterozygosity per se and concluded that vigor was correlated with the degree of heterozygosis. East (12) added to this explanation by assuming a series of divergent alleles at a locus where each member is dominant to its absence for a specific physiological process. By assuming that these effects are cumulative the effect of the heterozygotes was postulated to be greater than the homozygotes.

Many experiments have been conducted with the specific aim of

determining the relative importance of the types of gene action postulated under the two hypotheses. A complete review of the literature pertaining to the manifestation of heterosis and to the theories of heterosis has been presented by Sprague (55) and no attempt will be made here to present the data in support of either hypothesis. No evidence is available to indicate either hypothesis exists in the absence of the other and further studies are necessary to indicate the relative importance of each.

Comstock and Robinson (7) presented a method of obtaining the average degree of dominance under the assumption of no epistacy, no linkage, and gene frequency of one-half. From the analysis of variance of biparental progenies, dominance was calculated as the square root of twice the ratio of dominance variance to additive genetic variance. Robinson, et al. (47) obtained estimates from biparental progenies of three single cross corn populations grown in a single test in a single year. The estimate of average dominance for yield was " $a$ " = 1.64. If dominance is complete, " $a$ " = 1.0, and values of " $a$ " > 1 may indicate overdominance. An alternative explanation given was that as large a value of " $a$ " may result from tight linkages of certain genes in the repulsion phase, even though no individual genes may have more than partial dominance.

Garner, et al. (16) obtained estimates of dominance from random  $F_2$  plants crossed back to each parent for two populations. Estimates were obtained at two locations in two years. For yield, the " $a$ " values were again larger than 1.0. Horner (23) examined the procedure for bias from

epistacy and found no consistent bias from any of the gene interaction systems he considered. Absence of bias does not indicate absence of epistacy, but does indicate that the estimate of dominance variance should be free from it. Linkage is given as a possible important bias in the estimates of " $a$ ".

Robinson, et al. (48) obtained estimates of additive genetic variance and dominance variance from three open-pollinated varieties of corn by the biparental method. Epistacy, linkage, and genotype-environmental interaction were considered only in terms of possible bias they would give to the estimates of the additive and dominance variances. Estimates of additive genetic variance were considerably larger than those of dominance variance, and in some cases the magnitude of the additive genetic variance was many times larger than that of the dominance variance. The relative low value of dominance variance indicated that overdominance was not the only major source of genetic variability present in the material. The two possible reasons suggested for the ineffectiveness of intra-variety selection from open-pollinated material were:

1. There may be a negative correlation between grain yield and net reproductive capacity

2. The additive genetic variance may arise from loci where the gene action is additive, and where gene frequency is at equilibrium between mutation and selection forces.



### Combining Ability

Methods of testing lines of corn depend upon the assumptions one makes concerning gene action. Sprague and Tatum (57) compared the relative importance of general and specific combining ability in corn. General combining ability of a line is the average value of the line in all other combinations. The ability of a line to do better or worse than its average value in a specific cross is called specific combining ability. In highly selected material the variance of specific combining ability was larger than the variance of general combining ability, while for unselected lines the estimates of the variance of general combining ability were the larger. General combining ability was assumed to be a measure of additive gene action, while specific combining ability measured deviations from additivity.

Studies by Green (17), Keller (32), Matzinger (42), and Federer and Sprague (14) have given evidence of line x tester interactions in corn. Sprague and Federer (56) obtained variety x district and variety x year interactions with the variety x year interaction being of greatest magnitude.

Rojas (49) presented models to obtain estimates of variance components for general combining ability ( $\sigma_g^2$ ) and specific combining ability ( $\sigma_s^2$ ) and their interactions with years. Rojas and Sprague (50) obtained estimates of  $\sigma_s^2$  and  $\sigma_g^2$  over several locations for a three year period. Interaction components involving  $\sigma_s^2$  were consistently larger than those for  $\sigma_g^2$  indicating that genotype-environmental interaction may be part of

the measure of non-additivity in estimates of  $\sigma_s^2$ . The material in the study had been subjected to previous selection, which would tend to reduce the value of  $\sigma_g^2$  in relation to  $\sigma_s^2$  and possibly also the interactions with  $\sigma_g^2$ . In this experiment, also, interactions of both  $\sigma_s^2$  and  $\sigma_g^2$  were larger with years than with locations.

The ratio of genotypic variance to phenotypic variance is known as heritability in the broad sense. Heritability in the narrow sense is the ratio of additive genetic variance to phenotypic variance. A large number of experiments have been conducted to obtain estimates of heritability in a wide range of crops. Most of these have been estimates in the broad sense, and are reviewed by Byrd (4).

#### Analysis of Diallel Table

The analysis of variance technique was proposed by Yates (64) for the analysis of diallel crosses for the cases of self-sterility, no self-sterility, and self-sterility with incompatibility within a group of lines. Rojas (49) presented least squares estimates for the analysis of variance of diallel crosses and for the estimation of components of variance of diallel crosses in different locations and years in terms of the variance of specific and general combining ability.

Hayman (20,21) included dominance in the diallel cross analysis and considered a genetic system with the following assumptions:

1. Diploid segregation

2. No difference between reciprocal crosses
3. Independent action of non-allelic genes
4. No multiple allelism
5. Homozygous parents
6. Genes independently distributed between the parents.

Jinks (28) applied the method of analysis proposed by Hayman to the study of height, flowering time, and leaf length in Nicotiana rustica. Estimates of D and  $H_1$ , weighted values of Mather's (40)  $\sum d^2$  and  $\sum h^2$ , were used to calculate average degrees of dominance by the ratio  $H_1/D$ . The regression of array covariance on variance was plotted to obtain evidence of non-additive gene effects. Deviations from expected slope of unity indicated gene interaction might be responsible for observed heterosis of some  $F_1$  hybrids.

A summary of available data on the dialled cross was given by Jinks (29) for experiments on several important crop species, most of which had originally been conducted for another purpose. They include all possible intercrosses of from three to ten parents. By plotting the regression of covariance arrays on variance arrays, a deviation from unit slope of less than unity was considered to be evidence of non-allelic interaction, independent of degree of dominance. Lines which exhibited this interaction in crosses were removed from the analysis, and the remaining material was used to estimate the degree of dominance. Estimates of dominance values suggesting overdominance were reduced in all cases when the lines which exhibited non-allelic interaction were removed from the analysis.

Dickinson and Jinks (10) extended the results of Hayman and Jinks to include inbreeding of the parental material in the diallel cross. Beginning with an arbitrary group of parents, not necessarily representing a sample from a random mating population, estimates of correlations between relatives were obtained. Methods were given to compute the inbreeding coefficient during the course of the experiment.

Kempthorne (37) removed the assumptions (3) and (4) of Hayman and presented the analysis of the diallel cross for arbitrary alleles, arbitrary loci, and arbitrary epistacy. The parents in the diallel table were homozygous lines obtained by selfing without selection from a random mating population. Estimates of the mean and variance of both the random mating and homozygous populations, and the parent-offspring covariances were obtained from the variances and covariances of the diallel table. Evidence was given to show that previous methods of analysis of diallel crosses are of questionable validity unless one can assume no epistacy.

In the present thesis the results of Kempthorne (37) will be extended to include the case of arbitrary inbreeding for the parents of the diallel cross table.

## THEORETICAL DERIVATION OF COVARIANCES

The partition of the genotypic variance of inbred populations is considered in relation to genetic parameters of a random mating population. Since random mating is the simplest of all mating systems it has been possible to determine what happens to the additive genetic variance, dominance variance, and epistatic variance under this type of breeding system.

Let  $P$  be the phenotypic value of an individual,  $G$  be the genotypic value of an individual, and  $E$  be the contribution of environment. If the effects of the genotype and environment are additive,

$$P = G + E$$

where the genotypic value is considered to be an average of the phenotypes over the population of environments. Assuming that there is no correlation between the genotype and environment, the phenotypic variance is expressed as

$$\sigma_P^2 = \sigma_G^2 + \sigma_E^2$$

where  $\sigma_P^2$  is the phenotypic variance,  $\sigma_G^2$  the genotypic variance, and  $\sigma_E^2$  the environmental variance.

As a method of determining what happens to the additive, dominance, and epistatic deviations under a random mating system when individuals

are inbred and then crossed in all possible crosses, the theoretical covariances are calculated for these groups of deviations. A discussion is given for the cases of a single locus with two alleles and a single locus with multiple alleles to give a framework from which to develop a general case for two loci. Throughout the derivation presented here, the results are compared to other estimates in the literature by making comparable assumptions.

#### Single Locus with Two Alleles

In a random mating population at equilibrium the genotypic array for a single locus, A, with two alleles is

$$p^2 AA + 2pqAa + q^2 aa$$

and the gametic array is

$$pA + qa$$

where p is the gene frequency of A, q is the frequency of a, and

$$p + q = 1.$$

If two individuals are mated which are more closely related than the average of the population, the resulting progeny are inbred. The two genes at a locus of an individual can be alike because they descend from one gene or because of random sampling from a population. The probability of two genes being identical by descent is F, the inbreeding coefficient

which varies in the range from zero in a random mating population to one in an inbred population. Inbreeding and gene frequency are completely independent, with inbreeding changing the association of the alleles in pairs while gene frequency is the proportion of each allele in the population. Under inbreeding the genotypic array is

$$(p^2 + Fpq)AA + 2(1 - F)pqAa + (q^2 + Fpq)aa. \quad (1)$$

Table 1. Genotypic frequencies and coded values for a single locus with two alleles and inbreeding

Genotype	Frequency	Coded frequency	Genotypic value	Coded genotypic value
AA	$p^2 + Fpq$	$P_{22}$	d	y
Aa	$2(1 - F)pq$	$P_{21}$	h	$\frac{y - x}{2}$
aa	$q^2 + Fpq$	$P_{11}$	r	0

For simplicity the frequencies and the genotypic values are given a code shown in Table 1.

A mating of two individuals drawn at random from a random mating population with inbreeding and no selection had a mating structure as shown in Table 2.

From the offspring arrays the following variances were obtained:

1. Variance within families

Table 2. Mating structure of two individuals at one locus with two alleles for all mating types

Mating type	Sire genotype	Dam genotype	Frequency	Offspring array
1	AA	AA	$P_{22}^2$	AA
2	AA	Aa	$P_{22}P_{21}$	$\frac{1}{2}AA + \frac{1}{2}Aa$
3	AA	aa	$P_{22}P_{11}$	Aa
4	Aa	AA	$P_{21}P_{22}$	$\frac{1}{2}AA + \frac{1}{2}Aa$
5	Aa	Aa	$P_{21}^2$	$\frac{1}{4}AA + \frac{1}{2}Aa + \frac{1}{4}aa$
6	Aa	aa	$P_{21}P_{11}$	$\frac{1}{2}Aa + \frac{1}{2}aa$
7	aa	AA	$P_{11}P_{22}$	Aa
8	aa	Aa	$P_{11}P_{21}$	$\frac{1}{2}Aa + \frac{1}{2}aa$
9	aa	aa	$P_{11}^2$	aa

2. Variance between family means

3. Variance within sires

4. Variance between sire means

5. Total genotypic variance.

#### Variance within families

Mating types 1, 3, 7, and 9 will have no variance within families since all individuals within a family are genetically identical. The within variance obtained from Tables 1 and 2 by the usual procedure of



obtaining variances is

$$\begin{aligned}
 & 2P_{22}P_{21} \left[ \frac{1}{2}y^2 + \frac{1}{2}\left(\frac{y-x}{2}\right)^2 - \left\{ \frac{1}{2}y + \frac{1}{2}\left(\frac{y-x}{2}\right) \right\}^2 \right] \\
 & + P_{21}^2 \left[ \frac{1}{4}y^2 + \frac{1}{2}\left(\frac{y-x}{2}\right)^2 - \left\{ \frac{1}{4}y + \frac{1}{2}\left(\frac{y-x}{2}\right) \right\}^2 \right] \\
 & + 2P_{21}P_{11} \left[ \frac{1}{2}\left(\frac{y-x}{2}\right)^2 - \left\{ \frac{1}{2}\left(\frac{y-x}{2}\right) \right\}^2 \right] .
 \end{aligned}$$

After substituting for the  $P_{ij}$ 's and collecting terms the simplified expression is

$$\frac{1}{4}(1-F)pq \left[ y^2 + \left\{ 1 - (1-F)pq \right\} x^2 + 2(p-q)xy \right] . \quad (2)$$

If the original two lines were homozygous,  $F = 1$ , the above formula equals zero. This merely verifies that the  $F_1$  of a cross of two inbred lines has no genetic segregation.

As a check on the formula for random mating populations,  $F = 0$ , the coded genotypic values are transformed to the notation of Mather (40). Mather lets  $d = d$ ,  $h = h$ , and  $r = -d$ , where  $y = d - r$  and  $x = d + r - 2h$ . Considering his specific case of  $p = q = \frac{1}{2}$ , formula (2) becomes

$$\frac{1}{4}d^2 + \frac{3}{16}h^2$$

and by definition  $\sum d^2 = D$ , and  $\sum h^2 = H$ , giving the equation in its simplest form as

$$\frac{1}{4}D + \frac{3}{16}H. \quad (3)$$

Kempthorne (35) showed that for a random mating population the expected variance for within families is

$$\sigma_P^2 - \text{Cov}(\text{FS}) \quad (4)$$

where  $\text{Cov}(\text{FS})$  is the covariance of full-sibs and  $\sigma_P^2$  is the phenotypic variance. By definition

$$\sigma_P^2 = \sigma_A^2 + \sigma_D^2$$

for a single locus, where  $\sigma_A^2$  is the additive genetic variance and  $\sigma_D^2$  is the dominance variance. From the derivation of the covariance of relatives given by Kempthorne (35),

$$\text{Cov}(\text{FS}) = \frac{1}{2} \sigma_A^2 + \frac{1}{4} \sigma_D^2$$

and the expected variance for within families by substitution in (4) is

$$\frac{1}{2} \sigma_A^2 + \frac{3}{4} \sigma_D^2. \quad (5)$$

Kempthorne (38) defined additive genetic variance as the sum of squares which can be attributed to a regression of the phenotype of the individual on the number of genes of each possible type which the individual possesses. In the case of two alleles it is the regression of the values  $d$ ,  $h$ , and  $r$  on the number of  $A$  genes, and following the notation of Table 1 is

$$\sigma_A^2 = 2pq \left\{ p(d - h) + q(h - r) \right\}^2.$$

For a single locus the dominance variance is the difference between the total genotypic variance ( $\sigma_G^2$ ) and  $\sigma_A^2$ , which is expressed as

$$\sigma_D^2 = p^2 q^2 \{d + r - 2h\}^2.$$

If  $p = q = \frac{1}{2}$ , as in Mather (40),

$$\sigma_A^2 = \frac{1}{2}d^2 \quad (6)$$

and

$$\sigma_D^2 = \frac{1}{4}h^2. \quad (7)$$

Substituting for  $\sigma_A^2$  and  $\sigma_D^2$  in (5), the expected variance for within families is

$$\frac{1}{4}D + \frac{3}{16}H$$

which checks with equation (3).

#### Variance between family means

It will be seen that this computation is in fact unnecessary but it is included to give the reader a further indication of procedure. The variance between family means can be obtained directly from Table 2 using the coded genotypic values from Table 1 as

$$\begin{aligned} & p_{22}^2 y^2 + p_{22}p_{21} \left\{ \frac{1}{2}y + \frac{1}{2}\left(\frac{y-x}{2}\right) \right\}^2 + p_{22}p_{11} \left(\frac{y-x}{2}\right)^2 + p_{21}p_{22} \left\{ \frac{1}{2}y + \frac{1}{2}\left(\frac{y-x}{2}\right) \right\}^2 \\ & + p_{21}^2 \left\{ \frac{1}{4}y + \frac{1}{2}\left(\frac{y-x}{2}\right) \right\}^2 + p_{21}p_{11} \left\{ \frac{1}{2}\left(\frac{y-x}{2}\right) \right\}^2 + p_{11}p_{22} \left(\frac{y-x}{2}\right)^2 \end{aligned}$$

$$\begin{aligned}
& + P_{11}P_{21} \left\{ \frac{1}{2} \left( \frac{y-x}{2} \right) \right\}^2 - \left[ P_{22}^2 y + P_{22}P_{21} \left\{ \frac{1}{2}y + \frac{1}{2} \left( \frac{y-x}{2} \right) \right\} + P_{22}P_{11} \left( \frac{y-x}{2} \right) \right. \\
& + P_{21}P_{22} \left\{ \frac{1}{2}y + \frac{1}{2} \left( \frac{y-x}{2} \right) \right\} + P_{21}^2 \left\{ \frac{1}{4}y + \frac{1}{2} \left( \frac{y-x}{2} \right) \right\} + P_{21}P_{11} \left\{ \frac{1}{2} \left( \frac{y-x}{2} \right) \right\} \\
& \left. + P_{11}P_{22} \left( \frac{y-x}{2} \right) + P_{11}P_{21} \left\{ \frac{1}{2} \left( \frac{y-x}{2} \right) \right\} \right]^2 .
\end{aligned}$$

After substitution of the frequencies from Table 1 and combining of terms the simplified expression is

$$\frac{1}{4}(1+F)pq \left[ y^2 + \left\{ 1 - (3-F)pq \right\} x^2 + 2(p-q)xy \right] . \quad (8)$$

#### Variance within sires

To obtain the variance within sires over all dams, a new mating structure is given in Table 3. Since the individuals are drawn at random from a population, the female gametes represent the gametic array of the original random mating population, and each of the three possible males mate with these females in the frequencies indicated.

The variance within sires, from Table 3, using the same genotypic values as previously is,

$$\begin{aligned}
& P_{22} \left[ py^2 + q \left( \frac{y-x}{2} \right)^2 - \left\{ py + q \left( \frac{y-x}{2} \right) \right\}^2 \right] \\
& + P_{21} \left[ \frac{1}{2}py^2 + \frac{1}{2}q \left( \frac{y-x}{2} \right)^2 + \frac{1}{2}p \left( \frac{y-x}{2} \right)^2 - \left\{ \frac{1}{2}py + \frac{1}{2}q \left( \frac{y-x}{2} \right) + \frac{1}{2}p \left( \frac{y-x}{2} \right) \right\}^2 \right] \\
& + P_{11} \left[ p \left( \frac{y-x}{2} \right) - \left\{ p \left( \frac{y-x}{2} \right) \right\}^2 \right]
\end{aligned}$$

which simplifies to

Table 3. Mating structure of two individuals considering one locus with two alleles for sire mating types only

Sire	Frequency	Dam gametes	Offspring array
AA	$P_{22}$	$pA + qa$	$pAA + qAa$
Aa	$P_{21}$	$pA + qa$	$\frac{1}{2}(pAA + qAa)$ $+ \frac{1}{2}(pAa + qaA)$
aa	$P_{11}$	$pA + qa$	$pAa + qaA$

$$\frac{pq}{4} \left[ \frac{1}{2}(3 - F)y^2 + \left\{ 1 + \left( \frac{1}{2} - 2pq \right)(1 - F) \right\} x^2 + (3 - F)(p - q)xy \right]. \quad (9)$$

A check of this formula can be made for random mating populations for  $F = 0$ , and with  $p = q = 0.5$ . Expression (9) reduces to

$$\frac{3}{8} D + \frac{1}{4} H. \quad (10)$$

Kempthorne (35) derived the within sire variance for a random mating population as

$$\sigma_P^2 - \text{Cov}(\text{HS}) \quad (11)$$

where  $\text{Cov}(\text{HS})$  is the covariance of half-sibs. Using the method of Kempthorne (35) the  $\text{Cov}(\text{HS})$  is found to be  $\frac{1}{4} \sigma_A^2$ . Since

$$\sigma_P^2 = \sigma_A^2 + \sigma_D^2,$$

equation (11) becomes

$$\frac{3}{4} \sigma_A^2 + \sigma_D^2.$$

Substituting for  $\sigma_A^2$  and  $\sigma_D^2$  from equations (6) and (7) gives

$$\frac{3}{8} D + \frac{1}{4} H$$

which checks with equation (10).

#### Variance between sire means

The between sire variance can be written from Table 3 as

$$P_{22} \left\{ py + q \left( \frac{y-x}{2} \right) \right\}^2 + P_{21} \left\{ \frac{1}{2} py + \frac{1}{2} \left( \frac{y-x}{2} \right) \right\}^2 + P_{11} \left\{ p \left( \frac{y-x}{2} \right) \right\}^2 \\ - \left[ P_{22} \left\{ py + q \left( \frac{y-x}{2} \right) \right\} + P_{21} \left\{ \frac{1}{2} py + \frac{1}{2} \left( \frac{y-x}{2} \right) \right\} + P_{11} \left\{ p \left( \frac{y-x}{2} \right) \right\} \right]^2$$

and after simplification is

$$pq(1+F) \left\{ \frac{1}{8} y^2 + \left( \frac{1}{8} - \frac{1}{2} pq \right) x^2 + \frac{1}{4} (p-q) xy \right\}. \quad (12)$$

As a check of this formula for random mating,  $F = 0$ , and with  $p = \frac{1}{2}$ , the above equation becomes  $\frac{1}{8} D$ . In a random mating population the variance would be the covariance of half-sibs which was shown earlier to be  $\frac{1}{4} \sigma_A^2$ . Substitution of  $\frac{1}{2} d^2$  for  $\sigma_A^2$  gives the value of  $\frac{1}{8} D$  which checks with the general formula.

#### Total genotypic variance

The total genotypic variance can be computed directly from Table 1 as

$$(p^2 + Fpq)y^2 + 2(1 - F)pq\left(\frac{y - x}{2}\right)^2 - \left\{(p^2 + Fpq)y + 2(1 - F)pq\left(\frac{y - x}{2}\right)\right\}^2$$

and simplifies to

$$pq \left\{ \frac{1}{2}(1 + F)y^2 - (1 - F)^2 pqx^2 + \frac{1}{2}(1 - F)x^2 + (1 - F)(p - q)xy \right\} \quad (13)$$

as given in Kempthorne (36).

For a random mating population the total genotypic variance is

$$\frac{1}{2}pqy^2 - p^2q^2x^2 + \frac{1}{2}pqx^2 - pq(p - q)xy \quad (14)$$

after substitution for  $F = 0$  in equation (13).

As a check on the variances derived in this section, the variance between family means (8) plus the variance within families (2), and the variance between sire means (12) plus the variance within sires (9) equal the total genotypic variance for a random mating population (14).

The variances of this section serve as an aid in the details of the general case to be considered later and will also serve as a check of the more complex variances.

#### Single Locus, Arbitrary Alleles

This section contains the extension of the genetic model from two alleles per locus to any number of alleles. With multiple alleles  $A_1, A_2, \dots, A_m$ , with corresponding frequencies  $p_1, p_2, \dots, p_m$  where  $\sum_1 p_i = 1$ , the genotypic array in a random mating population is

$$\left\{ \sum_{i=1}^m p_i A_i \right\}^2 = \sum_i p_i^2 A_i A_i + 2 \sum_{\substack{i,j \\ i < j}} p_i p_j A_i A_j$$

and in a population inbred to  $F$  is

$$\sum_i \left\{ (1-F)p_i^2 + Fp_i \right\} A_i A_i + 2(1-F) \sum_{\substack{i,j \\ i < j}} p_i p_j A_i A_j$$

which can also be written as

$$F \sum_i p_i A_i A_i + (1-F) \sum_{i,j} p_i p_j A_i A_j .$$

By summing on  $i$  and  $j$  for values of  $m = 1$  and  $2$ , this genotypic array simplifies to that of a population with only two alleles per locus, equation (1).

Let  $P_{ij}$  be the frequency of individual  $A_i A_j$  where

$$P_{ij} = (1-F)p_i p_j + \delta_{ij} F p_i$$

and

$$\begin{aligned} \delta_{ij} &= 1 \text{ if } i = j \\ &= 0 \text{ if } i \neq j . \end{aligned}$$

Then

$$\sum_j P_{ij} = p_i .$$

Kempthorne (34,35) gave definitions of basic genetic components of



a random mating population and they will be used in the present derivations. Considering a single locus with multiple alleles the genotype of  $A_i A_j$  can be expressed as

$$y_{ij} = \mu + \alpha_i + \alpha_j + d_{ij}$$

where  $\alpha_i$  is the additive genetic value of the  $i$ -th allele and  $d_{ij}$  is the dominance deviation. The  $y_{ij}$ 's are measured from the population mean giving

$$\sum_i p_i \alpha_i = 0.$$

If we code so that the mean of the random mating population be zero, then the total genotypic variance is

$$\sigma_G^2 = \sum_{ij} p_i p_j y_{ij}^2.$$

the additive value at the  $i$ -th locus is

$$\alpha_i = \sum_j p_j y_{ij}.$$

the additive genetic variance is

$$\sigma_A^2 = 2 \sum_i p_i \alpha_i^2,$$

and the dominance variance is

$$\sigma_D^2 = \sum_{ij} p_i p_j d_{ij}^2.$$

### Variance between family means

The mating structure of a population considering a single locus with an arbitrary number of alleles is given in Table 4.

The mean of the offspring is

$$\frac{1}{4}(y_{ik} + y_{il} + y_{jk} + y_{jl})$$

Table 4. Mating structure of family means for one locus with arbitrary alleles

Sire	Frequency	Dam	Frequency	Offspring array
$A_i A_j$	$P_{ij}$	$A_k A_l$	$P_{kl}$	$\frac{1}{4} A_i A_k + \frac{1}{4} A_i A_l + \frac{1}{4} A_j A_k + \frac{1}{4} A_j A_l$

and the variance between family means can be expressed as

$$\frac{1}{16} \sum_{ijkl} P_{ij} P_{kl} (y_{ik} + y_{il} + y_{jk} + y_{jl})^2$$

which after substituting in terms of additive and dominance effects is

$$\frac{1}{16} \sum_{ijkl} P_{ij} P_{kl} (2\alpha_i + 2\alpha_j + 2\alpha_k + 2\alpha_l + d_{ik} + d_{il} + d_{jk} + d_{jl})^2.$$

In the expansion of this variance the mean will be the mean of the random mating population which is defined to be zero. The evaluation

of the variance is the summation of the eight squared terms and the twenty-eight cross product terms. The expansion of the variance is simplified by use of definitions and symbols given previously.

The eight squared terms are of two basic types. There are four terms of the type  $\alpha_i^2$  which simplify as

$$\frac{1}{4} \sum_{ijkl} p_{ij} p_{kl} \alpha_i^2 = \frac{1}{4} \sum_i p_i \alpha_i^2 = \frac{1}{8} \sigma_A^2 .$$

The other four terms are of the type

$$\frac{1}{16} \sum_{ijkl} p_{ij} p_{kl} d_{ik}^2 = \frac{1}{16} \sum_{ik} p_i p_k d_{ik}^2 = \frac{1}{16} \sigma_D^2 .$$

The twenty-eight cross product terms are of five basic types. Two cross product terms between the  $\alpha$ 's have subscripts from only one parent as

$$\begin{aligned} \frac{1}{2} \sum_{ijkl} p_{ij} p_{kl} \alpha_i \alpha_j &= \frac{1}{2} \sum_{ij} p_{ij} \alpha_i \alpha_j \\ &= \frac{1}{2} \sum_{ij} \left\{ (1 - F) p_i p_j + \delta_{ij} F p_i \right\} \alpha_i \alpha_j \\ &= \frac{1}{2} F \sum_i p_i \alpha_i^2 \\ &= \frac{1}{4} F \sigma_A^2 . \end{aligned}$$

When the subscripts are from both parents as in four terms, the simplification is

$$\frac{1}{2} \sum_{ijkl} P_{ij} P_{kl} \alpha_i \alpha_k = \frac{1}{2} \sum_{ik} p_i p_k \alpha_i \alpha_k = 0.$$

Four of the cross product terms between the d's involve only three subscripts and can be simplified as

$$\begin{aligned} & \frac{1}{8} \sum_{ijkl} P_{ij} P_{kl} d_{ik} d_{jl} \\ &= \frac{1}{8} \sum_{ijkl} p_i \left\{ (1-F) p_k p_l + \delta_{kl} F p_k \right\} d_{ik} d_{jl} \\ &= \frac{1}{8} F \sum_{ik} p_i p_k d_{ik}^2 \\ &= \frac{1}{8} F \sigma_D^2. \end{aligned}$$

The other two cross product terms between the d's contain four different subscripts as

$$\begin{aligned} & \frac{1}{8} \sum_{ijkl} P_{ij} P_{kl} d_{ik} d_{jl} \\ &= \frac{1}{8} \sum_{ijkl} \left\{ (1-F) p_i p_j + \delta_{ij} F p_i \right\} \left\{ (1-F) p_k p_l + \delta_{kl} F p_k \right\} d_{ik} d_{jl} \\ &= \frac{1}{8} F^2 \sum_{ik} p_i p_k d_{ik}^2 \\ &= \frac{1}{8} F^2 \sigma_D^2. \end{aligned}$$

All sixteen of the cross product terms between the  $\alpha$ 's and d's equal zero as

$$\frac{1}{4} \sum_{ijkl} p_{ij} p_{kl} \alpha_{ij} d_{ik} = \frac{1}{4} \sum_{ik} p_i p_k \alpha_i d_{ik} = 0.$$

The variance between family means is then obtained by summation of these 36 terms. This sum is

$$\begin{aligned} & \frac{1}{2} \sigma_A^2 + \frac{1}{4} \sigma_D^2 + \frac{1}{2F} \sigma_A^2 + \frac{1}{2F} \sigma_D^2 + \frac{1}{4F^2} \sigma_D^2 \\ &= \left( \frac{1+F}{2} \right) \sigma_A^2 + \left( \frac{1+F}{2} \right)^2 \sigma_D^2. \end{aligned} \quad (15)$$

A check of this formula for the specific case of two alleles per locus can be made. Substitution in equation (15) for  $\sigma_D^2$  and  $\sigma_A^2$  for the case of two alleles per locus, where

$$\sigma_D^2 = p^2 q^2 x^2 \quad (16)$$

and

$$\sigma_A^2 = \frac{1}{2} p q \{ y + (p - q)x \}^2 \quad (17)$$

equation (15) is expressed as

$$\frac{1}{4} (1+F)^2 p^2 q^2 x^2 + \frac{1}{2} (1+F) \left[ \frac{1}{2} p q \{ y + (p - q)x \}^2 \right].$$

This reduces to

$$\frac{1}{4} (1+F) p q \left[ y^2 + \{ 1 - p q (3 - F) \} x^2 + 2(p - q)xy \right]$$

which checks with formula (8) for the case of two alleles per locus.

### Variance within families

The variance within families is obtained by subtraction of the variance between family means from the total genotypic variance. Thus

$$\left\{1 - \left(\frac{1+F}{2}\right)\right\} \sigma_A^2 + \left\{1 - \left(\frac{1+F}{2}\right)^2\right\} \sigma_D^2$$

reduces to

$$\frac{1}{2}(1-F) \sigma_A^2 + \frac{1}{4}(1-F)(3+F) \sigma_D^2. \quad (18)$$

Substitution for  $\sigma_D^2$  and  $\sigma_A^2$  for two alleles per locus from equations (16) and (17),

$$\begin{aligned} &= \frac{1}{2}(1-F) \left[ \frac{1}{2}pq \left\{ y + (p-q)x \right\}^2 \right] + \frac{1}{4}(1-F)(3+F)p^2q^2x^2 \\ &= \frac{1}{4}(1-F)pq \left[ y^2 + \left\{ 1 - (1-F)pq \right\} x^2 + 2(p-q)xy \right] \end{aligned}$$

which checks with equation (2) for the case of two alleles per locus.

### Variance within sires

The mating system for the sires is given in Table 5. The dam gametes represent the gametic array of the original population. Using the same symbols as in the previous section the mean of the offspring array is

$$\frac{1}{2}(\alpha_i + \alpha_j)$$

where  $\alpha_i$  and  $\alpha_j$  are measures of the additive gene effects of the  $i$ -th and  $j$ -th locus respectively. Then the variance within sires coding  $\mu_R$  as zero is

Table 5. Mating structure of two individuals considering multiple alleles at one locus for sire mating types

Sire	Frequency	Offspring array
$A_i A_j$	$P_{ij}$	$\frac{1}{2} \sum_s p_s A_i A_s + \frac{1}{2} \sum_s p_s A_j A_s$

$$\sum_{ij} P_{ij} \left\{ \frac{1}{2} \sum_s p_s y_{is}^2 + \frac{1}{2} \sum_s p_s y_{js}^2 - \left( \frac{1}{2} \alpha_i + \frac{1}{2} \alpha_j \right)^2 \right\}.$$

The first two terms simplify as

$$\begin{aligned} \frac{1}{2} \sum_{ij} P_{ij} \sum_s p_s y_{is}^2 &= \frac{1}{2} \sum_{is} p_i p_s y_{is}^2 \\ &= \frac{1}{2} \sigma_G^2. \end{aligned}$$

The two terms of  $\alpha$  when squared simplify to

$$\begin{aligned} \frac{1}{4} \sum_{ij} P_{ij} \alpha_i^2 &= \frac{1}{4} \sum_i p_i \alpha_i^2 \\ &= \frac{1}{8} \sigma_A^2. \end{aligned}$$

The cross product term of the  $\alpha$ 's is

$$\frac{1}{2} \sum_{ij} P_{ij} \alpha_i \alpha_j = \frac{1}{2} \sum_i p_i \alpha_i^2$$

$$= \frac{1}{4} F \sigma_A^2 .$$

After addition of these terms the variance within sires is

$$\frac{1}{4}(3 - F) \sigma_A^2 + \sigma_D^2 . \quad (19)$$

Substitution for  $\sigma_A^2$  and  $\sigma_D^2$  in relation to two alleles per locus gives

$$\frac{1}{4}(3 - F) \left[ \frac{1}{2}pq \left\{ y + (p - q)x \right\}^2 \right] + p^2 q^2 x^2$$

which simplifies to

$$\frac{pq}{4} \left[ \frac{1}{2}(3 - F)y^2 + \left\{ 1 + \left( \frac{1}{2} - 2pq \right)(1 - F) \right\} x^2 + (3 - F)(p - q)xy \right]$$

and checks with formula (9) for two alleles per locus.

#### Variance between sire means

The variance between sire means is obtained by the difference between total genotypic variance and the variance within sires. This difference is

$$\begin{aligned} \sigma_A^2 + \sigma_D^2 - \frac{1}{4}(3 - F) \sigma_A^2 - \sigma_D^2 \\ = \left( \frac{1 + F}{4} \right) \sigma_A^2 . \end{aligned} \quad (20)$$

Substituting for  $\sigma_A^2$  in the two allele case gives

$$\frac{1}{4}(1 + F) \left[ \frac{1}{2}pq \left\{ y + (p - q)x \right\}^2 \right]$$



$$= pq(1 + F) \left\{ \frac{1}{8}y^2 + \left( \frac{1}{8} - \frac{1}{2}pq \right)x^2 + \frac{1}{4}(p - q)xy \right\}$$

which checks with equation (12).

### Two Loci, Arbitrary Alleles

The results of the previous section are now outlined in detail for the case of two loci. The method is general for arbitrary alleles per locus and arbitrary epistacy and for an arbitrary number of loci by straight forward extension.

The genotypic value of an individual  $y_{ijkl}$  can be represented as

$$\begin{aligned} y_{ijkl} = & \mu + \alpha_i^1 + \alpha_j^1 + (\alpha^1\alpha^1)_{ij} + \alpha_k^2 + \alpha_l^2 + (\alpha^2\alpha^2)_{kl} \\ & + (\alpha^1\alpha^2)_{ik} + (\alpha^1\alpha^2)_{jk} + (\alpha^1\alpha^2)_{il} + (\alpha^1\alpha^2)_{jl} + (\alpha^1\alpha^2\alpha^2)_{ikl} \\ & + (\alpha^1\alpha^2\alpha^2)_{jkl} + (\alpha^1\alpha^1\alpha^2)_{ijk} + (\alpha^1\alpha^1\alpha^2)_{ijl} + (\alpha^1\alpha^1\alpha^2\alpha^2)_{ijkl} \end{aligned} \quad (21)$$

where the superscripts indicate the locus (Kempthorne, 38). For example,  $\alpha_i^1$  is the average effect of the  $i$ -th allele at locus 1,  $(\alpha^2\alpha^2)_{kl}$  is the dominance effect at locus 2,  $(\alpha^1\alpha^2)_{ik}$  is the interaction of the additive effect of the  $i$ -th allele at locus 1 with the additive effect of the  $k$ -th allele at locus 2, and the other terms have similar meanings. Note that the terminology for the dominance deviations has been changed, as for instance in the case of  $(\alpha^1\alpha^1)_{ij}$  which has previously been denoted by  $d_{ij}$ .

### Variance between family means

Consider a cross of  $A_1A_jB_kB_l \times A_mA_nB_rB_s$  where A and B are two independent loci, and the subscripts indicate the allele at that locus. Let the frequencies of  $A_1A_j$ ,  $B_kB_l$ ,  $A_mA_n$ , and  $B_rB_s$  be designated as  $P_{ij}$ ,  $P'_{kl}$ ,  $P_{mn}$ , and  $P'_{rs}$  respectively.

The variance between families is obtained as

$$\sum_{\substack{ijkl \\ mnrs}} P_{ij} P'_{kl} P_{mn} P'_{rs} \left[ \left\{ \frac{1}{2}(A_1 + A_j) \frac{1}{2}(A_m + A_n) \right\} \left\{ \frac{1}{2}(B_k + B_l) \frac{1}{2}(B_r + B_s) \right\} \right]^2$$

$$= \sum_{\substack{ijkl \\ mnrs}} P_{ij} P'_{kl} P_{mn} P'_{rs} \left\{ \left( \frac{1}{4}A_1A_m + \frac{1}{4}A_1A_n + \frac{1}{4}A_jA_m + \frac{1}{4}A_jA_n \right) \right.$$

$$\left. \left( \frac{1}{4}B_kB_r + \frac{1}{4}B_kB_s + \frac{1}{4}B_lB_r + \frac{1}{4}B_lB_s \right) \right\}^2$$

in which each genotype symbol is replaced by the corresponding genotypic value.

To solve this expression the terms inside of the brackets are written in the form of additive, dominance, and interaction effects of the type in equation (21). The formal expansion of this term is very detailed and is not written out completely. It is presented in tabular form in Table 6. To obtain the equation from this table, each subscript must be attached to the appropriate symbol, multiplied by the corresponding coefficient, and then summed over all eighty terms. The equation is then of the form

Table 6. Individual expressions of expansion of variance between families for two loci

Group of terms	Coefficient	Symbol	Subscripts			
A	1/2	$\alpha^1$	i	j	m	n
A	1/2	$\alpha^2$	k	l	r	s
D	1/4	$(\alpha^1 \alpha^1)$	im	in	jm	jn
D	1/4	$(\alpha^2 \alpha^2)$	kr	ks	lr	ls
A x A	1/4	$(\alpha^1 \alpha^2)$	ik	il	ir	is
			jk	jl	jr	js
			mk	ml	mr	ms
			nk	nl	nr	ns
A x D	1/8	$(\alpha^1 \alpha^2 \alpha^2)$	ikr	jkr	mkr	nkr
			iks	jks	mks	nks
			ilr	jl r	mlr	nlr
			ils	jls	mls	nls
D x A	1/8	$(\alpha^1 \alpha^1 \alpha^2)$	imk	iml	imr	ims
			ink	inl	inr	ins
			jm k	jml	jmr	jms
			jnk	jnl	jnr	jns
D x D	1/16	$(\alpha^1 \alpha^1 \alpha^2 \alpha^2)$	imkr	imks	imlr	imls
			inkr	inks	inlr	inls
			jmkr	jmks	jmlr	jmls
			jnkr	jnks	jnlr	jnls

$$\sum_{\substack{ijkl \\ mnrs}} P_{ij} P_{kl}^i P_{mn} P_{rs}^i (80 \text{ terms})^2 .$$

A detailed solution of this equation could be obtained by performing the square of the 80 terms, giving 3240 terms, and then completing the summation as indicated term by term. This is very tedious so only examples of each of the different terms are computed and then weighted by the number of times that type of term appears.

There are 80 squared terms. Since  $\sum_{ij} P_{ij} = 1$ , any of the  $P$ 's which have subscripts not on the particular term being evaluated do not enter into the computation. For example,

$$\sum_{\substack{ijkl \\ mnrs}} P_{ij} P_{kl}^i P_{mn} P_{rs}^i y_{ik} = \sum_{ijkl} P_{ij} P_{kl}^i y_{ik} .$$

The evaluation of only the first of each type of term is given. The total variance components are

$$\sigma_A^2 = \sigma_{A_1}^2 + \sigma_{A_2}^2$$

$$\sigma_D^2 = \sigma_{D_1}^2 + \sigma_{D_2}^2$$

$$\sigma_{AD}^2 = \sigma_{A_1 D_2}^2 + \sigma_{A_2 D_1}^2$$

where

$$\sigma_{A_1}^2 = \text{additive genetic variance from locus 1}$$

$\sigma^2_{A_2}$  = additive genetic variance from locus 2

$\sigma^2_{D_1}$  = dominance variance at locus 1

and the others follow in a similar manner.

The evaluation of the squared terms is as follows:

(a) Additive

$$\begin{aligned} \frac{1}{4} \sum_{ij} p_{ij} (\alpha_i^1)^2 &= \frac{1}{4} \sum_i p_i (\alpha_i^1)^2 \\ &= \frac{1}{8} \sigma^2_{A_1} \end{aligned}$$

(b) Dominance

$$\begin{aligned} \frac{1}{16} \sum_{ijmn} p_{ij} p_{mn} \left\{ (\alpha_i^1 \alpha_j^1)_{im} \right\}^2 &= \frac{1}{16} \sum_{im} p_i p_m \left\{ (\alpha_i^1 \alpha_j^1)_{im} \right\}^2 \\ &= \frac{1}{16} \sigma^2_{D_1} \end{aligned}$$

(c) Additive x dominance

$$\begin{aligned} \frac{1}{64} \sum_{\substack{ijkl \\ mn}} p_{ij} p'_{kl} p_{mn} \left\{ (\alpha_i^1 \alpha_j^1 \alpha_k^2)_{imk} \right\}^2 \\ &= \frac{1}{64} \sum_{imk} p_i p_m p'_k \left\{ (\alpha_i^1 \alpha_j^1 \alpha_k^2)_{imk} \right\}^2 \\ &= \frac{1}{128} \sigma^2_{A_2 D_1} \end{aligned}$$

(d) Additive x additive

$$\begin{aligned}
 & \frac{1}{16} \sum_{ijkl} p_{ij} p'_{kl} \left\{ (\alpha^1 \alpha^2)_{ik} \right\}^2 \\
 &= \frac{1}{16} \sum_{ik} p_i p'_k \left\{ (\alpha^1 \alpha^2)_{ik} \right\}^2 \\
 &= \frac{1}{64} \sigma^2_{A_1 A_2}
 \end{aligned}$$

(e) Dominance x dominance

$$\begin{aligned}
 & \frac{1}{256} \sum_{ijkl} p_{ij} p'_{kl} p_{mn} p'_{rs} \left\{ (\alpha^1 \alpha^1 \alpha^2 \alpha^2)_{imkr} \right\}^2 \\
 &= \frac{1}{256} \sum_{imkr} p_i p_m p'_k p'_r \left\{ (\alpha^1 \alpha^1 \alpha^2 \alpha^2)_{imkr} \right\}^2 \\
 &= \frac{1}{256} \sigma^2_{D_1 D_2}
 \end{aligned}$$

The cross product terms are considered by groups as follows:

(a) Within additive group of terms

(1) 1 locus, 1 parent

$$\begin{aligned}
 & \frac{1}{2} \sum_{ij} p_{ij} \alpha^1_i \alpha^1_j \\
 &= \frac{1}{2} \sum_{ij} \alpha^1_i \alpha^1_j \left\{ (1 - F) p_i p_j + \delta_{ij} F p_i \right\}
 \end{aligned}$$

$$= \frac{1}{2} F \sum_i p_i (\alpha_i^1)^2 = \frac{1}{4} F \sigma^2 A_1$$

(2) 1 locus, 2 parents

$$\begin{aligned} & \frac{1}{2} \sum_{ijmn} p_{ij} p_{mn} \alpha_i^1 \alpha_m^1 \\ &= \frac{1}{2} \sum_{im} p_i p_m \alpha_i^1 \alpha_m^1 \\ &= \frac{1}{2} \left( \sum_i p_i \alpha_i^1 \right)^2 = 0 \end{aligned}$$

(3) 2 loci

$$\begin{aligned} & \frac{1}{2} \sum_{ijkl} p_{ij} p_{kl} \alpha_i^1 \alpha_k^2 \\ &= \frac{1}{2} \sum_{ik} p_i p_k \alpha_i^1 \alpha_k^2 \\ &= \frac{1}{2} \left( \sum_i p_i \alpha_i^1 \right)^2 = 0 \end{aligned}$$

(b) Within dominance group of terms

(1) 3 different subscripts, 1 locus

$$\begin{aligned} & \frac{1}{8} \sum_{ijmn} p_{ij} p_{mn} (\alpha_i^1 \alpha_j^1)_{im} (\alpha_m^1 \alpha_n^1)_{in} \\ &= \frac{1}{8} \sum_{imn} p_i (\alpha_i^1 \alpha_j^1)_{im} (\alpha_m^1 \alpha_n^1)_{in} \left\{ (1-F) p_m p_n + \delta_{mn} F p_m \right\} \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{8}(1-F) \sum_{imn} p_i p_m p_n (\alpha^1 \alpha^1)_{im} (\alpha^1 \alpha^1)_{in} + \frac{1}{8}F \sum_{imn} p_i p_m \delta_{mn} (\alpha^1 \alpha^1)_{im} (\alpha^1 \alpha^1)_{in} \\
&= \frac{1}{8}(1-F) \sum_{im} p_i p_m (\alpha^1 \alpha^1)_{im} \sum_n p_n (\alpha^1 \alpha^1)_{in} + \frac{1}{8}F \sum_{im} p_i p_m \left\{ (\alpha^1 \alpha^1)_{im} \right\}^2 \\
&= \frac{1}{8}F \sigma_{D_1}^2
\end{aligned}$$

(2) 4 different subscripts, 1 locus

$$\begin{aligned}
&\frac{1}{8} \sum_{ijmn} p_{ij} p_{mn} (\alpha^1 \alpha^1)_{im} (\alpha^1 \alpha^1)_{jn} \\
&= \frac{1}{8} \sum_{ijmn} \left\{ (1-F)p_i p_j + \delta_{ij} F p_i \right\} \left\{ (1-F)p_m p_n + \delta_{mn} F p_m \right\} (\alpha^1 \alpha^1)_{im} (\alpha^1 \alpha^1)_{jn} \\
&= \frac{1}{8}F^2 \sum_{im} p_i p_m \left\{ (\alpha^1 \alpha^1)_{im} \right\}^2 \\
&= \frac{1}{8}F^2 \sigma_{D_1}^2
\end{aligned}$$

(3) 4 different subscripts, 2 loci

$$\begin{aligned}
&\frac{1}{8} \sum_{\substack{ijkl \\ mnrs}} p_{ij} p'_{kl} p_{mn} p'_{rs} (\alpha^1 \alpha^1)_{im} (\alpha^2 \alpha^2)_{kr} \\
&= \frac{1}{8} \sum_{ikmr} p_i p'_{k'} p_m p'_{r'} (\alpha^1 \alpha^1)_{im} (\alpha^2 \alpha^2)_{kr} \\
&= 0
\end{aligned}$$



## (c) Between additive and dominance

## (1) additive and dominance at different loci

$$\begin{aligned}
& \frac{1}{4} \sum_{\substack{ijkl \\ rs}} p_{ij} p_{kl} p_{rs} \alpha_i^1 (\alpha^2 \alpha^2)_{kr} \\
&= \frac{1}{4} \sum_{ikr} p_i p_k p_r \alpha_i^1 (\alpha^2 \alpha^2)_{kr} \\
&= 0
\end{aligned}$$

## (2) additive and dominance at same locus

$$\begin{aligned}
& \frac{1}{4} \sum_{ijmn} p_{ij} p_{mn} \alpha_i^1 (\alpha^1 \alpha^1)_{im} \\
&= \frac{1}{4} \sum_i p_i \alpha_i^1 \sum_m p_m (\alpha^1 \alpha^1)_{im} \\
&= 0
\end{aligned}$$

## (d) Within dominance x additive

## (1) 4 different subscripts, additive in 1 parent

$$\begin{aligned}
& \frac{1}{32} \sum_{\substack{ijkl \\ mn}} p_{ij} p_{kl} p_{mn} (\alpha^1 \alpha^1 \alpha^2)_{imk} (\alpha^1 \alpha^1 \alpha^2)_{iml} \\
&= \frac{1}{32} \sum_{imkl} p_i p_m (\alpha^1 \alpha^1 \alpha^2)_{imk} (\alpha^1 \alpha^1 \alpha^2)_{iml} \left\{ (1 - F) p_k^1 p_l^1 + \delta_{kl} F p_k^1 \right\}
\end{aligned}$$

$$= \frac{1}{32} F \sum_{imk} p_i p_m p'_k \left\{ (\alpha^1 \alpha^1 \alpha^2)_{imk} \right\}^2$$

$$= \frac{1}{64} F \sigma^2_{D_1 A_2}$$

(2) 4 different subscripts, additive in 2 parents

$$\frac{1}{32} \sum_{\substack{ijkl \\ mnrs}} p_i p'_j p'_k p'_l p_m p'_n p'_r p'_s (\alpha^1 \alpha^1 \alpha^2)_{imk} (\alpha^1 \alpha^1 \alpha^2)_{imr}$$

$$= \frac{1}{32} \sum_{imkr} p_i p_m p'_k p'_r (\alpha^1 \alpha^1 \alpha^2)_{imk} (\alpha^1 \alpha^1 \alpha^2)_{imr}$$

$$= 0$$

(3) 5 different subscripts, additive in 1 parent

$$\frac{1}{32} \sum_{\substack{ijkl \\ mn}} p_i p'_j p'_k p'_l p_m (\alpha^1 \alpha^1 \alpha^2)_{imk} (\alpha^1 \alpha^1 \alpha^2)_{iml}$$

$$= \frac{1}{32} \sum_{\substack{ikl \\ mn}} p_i p'_k p'_l p_m (\alpha^1 \alpha^1 \alpha^2)_{imk} (\alpha^1 \alpha^1 \alpha^2)_{iml}$$

$$= \frac{1}{32} \sum_{\substack{ikl \\ mn}} p_i (\alpha^1 \alpha^1 \alpha^2)_{imk} (\alpha^1 \alpha^1 \alpha^2)_{iml} \left\{ (1-F) p'_k p'_l + \delta_{kl} F p'_k \right\} \left\{ (1-F) p_m p_n + \delta_{mn} F p_m \right\}$$

$$= \frac{1}{32} F^2 \sum_{imk} p_i p'_k p'_m \left\{ (\alpha^1 \alpha^1 \alpha^2)_{imk} \right\}^2$$

$$= \frac{1}{64} F^2 \sigma^2_{D_1 A_2}$$

(4) 5 different subscripts, additive in 2 parents

$$\begin{aligned}
 & \frac{1}{32} \sum_{\substack{ijkl \\ mnrs}} P_{ij} P_{kl} P_{mn} P_{rs} (\alpha^1 \alpha^1 \alpha^2)_{imk} (\alpha^1 \alpha^1 \alpha^2)_{inr} \\
 &= \frac{1}{32} \sum_{\substack{imn \\ kr}} P_{ij} P_{kl} P_{rs} (\alpha^1 \alpha^1 \alpha^2)_{imk} (\alpha^1 \alpha^1 \alpha^2)_{inr} \left\{ (1 - F) p_m p_n + \delta_{mn} F p_m \right\} \\
 &= 0
 \end{aligned}$$

(5) 6 different subscripts, additive in 1 parent

$$\begin{aligned}
 & \frac{1}{32} \sum_{\substack{ijkl \\ mn}} P_{ij} P_{kl} P_{mn} (\alpha^1 \alpha^1 \alpha^2)_{iml} (\alpha^1 \alpha^1 \alpha^2)_{jnk} \\
 &= \frac{1}{32} \sum_{\substack{ijkl \\ mn}} (\alpha^1 \alpha^1 \alpha^2)_{iml} (\alpha^1 \alpha^1 \alpha^2)_{jnk} \left\{ (1 - F) p_i p_j + \delta_{ij} F p_i \right\} \left\{ (1 - F) p_k p_l \right. \\
 & \quad \left. + \delta_{kl} F p_k \right\} \left\{ (1 - F) p_m p_n + \delta_{mn} F p_m \right\} \\
 &= \frac{1}{32} \sum_{ikm} P_{ij} P_{kl} P_{mn} F^3 \left\{ (\alpha^1 \alpha^1 \alpha^2)_{imk} \right\}^2 \\
 &= \frac{1}{64} F^3 \sigma_{D_1 A_2}^2
 \end{aligned}$$

(6) 6 different subscripts, additive in 2 parents

$$\frac{1}{32} \sum_{\substack{ijkl \\ mnrs}} P_{ij} P_{kl} P_{mn} P_{rs} (\alpha^1 \alpha^1 \alpha^2)_{imk} (\alpha^1 \alpha^1 \alpha^2)_{jnr}$$

$$\begin{aligned}
&= \frac{1}{32} \sum_{\substack{ijk \\ mnr}} p_i p_j p_k p_m p_r (\alpha^1 \alpha^1 \alpha^2)_{imk} (\alpha^1 \alpha^1 \alpha^2)_{jnr} \\
&= \frac{1}{32} \sum_{\substack{ijk \\ mnr}} p_i p_j p_k p_r (\alpha^1 \alpha^1 \alpha^2)_{imk} (\alpha^1 \alpha^1 \alpha^2)_{jnr} \left\{ (1-F)p_i p_j + \delta_{ij} F p_i \right\} \left\{ (1-F)p_m p_n + \delta_{mn} F p_m \right\} \\
&= \frac{1}{32} F^2 \sum_{ikmr} p_i p_m p_k p_r (\alpha^1 \alpha^1 \alpha^2)_{imk} (\alpha^1 \alpha^1 \alpha^2)_{imr} \\
&= 0
\end{aligned}$$

(e) Within additive by dominance

These terms are the same as the within dominance by additive except that the dominance is now at locus 2 and the additive at locus 1.

(f) Between dominance x additive and additive x dominance

(1) 4 different subscripts

$$\begin{aligned}
&\frac{1}{32} \sum_{\substack{ijkl \\ mnr}} p_i p_j p_k p_l p_m p_r (\alpha^1 \alpha^1 \alpha^2)_{imk} (\alpha^1 \alpha^2 \alpha^2)_{mkr} \\
&= \frac{1}{32} \sum_{ikmr} p_i p_j p_k p_r (\alpha^1 \alpha^1 \alpha^2)_{imk} (\alpha^1 \alpha^2 \alpha^2)_{mkr} \\
&= 0
\end{aligned}$$

(2) 5 different subscripts

$$\frac{1}{32} \sum_{\substack{ijkl \\ mnr}} p_i p_j p_k p_l p_m p_r (\alpha^1 \alpha^1 \alpha^2)_{imk} (\alpha^1 \alpha^2 \alpha^2)_{jkr}$$

$$= \frac{1}{32} \sum_{\substack{ijk \\ nr}} p_i' p_k' p_m' p_r' \left\{ (1 - F) p_i p_j + \delta_{ij} F p_i \right\} (\alpha^1 \alpha^1 \alpha^2)_{imk} (\alpha^1 \alpha^2 \alpha^2)_{jkr}$$

$$= \frac{1}{32} \sum_{ikmr} p_i p_k' p_m' p_r' (\alpha^1 \alpha^1 \alpha^2)_{imk} (\alpha^1 \alpha^2 \alpha^2)_{ikr}$$

$$= 0$$

(3) 6 different subscripts

$$\frac{1}{32} \sum_{\substack{ijkl \\ mrs}} p_{ij} p_{kl}' p_{mn} p_{rs}' (\alpha^1 \alpha^1 \alpha^2)_{imk} (\alpha^1 \alpha^2 \alpha^2)_{jlr}$$

$$= \frac{1}{32} \sum_{\substack{ijk \\ lnr}} p_m' p_r' \left\{ (1 - F) p_i p_j + \delta_{ij} F p_i \right\} \left\{ (1 - F) p_k' p_l' + \delta_{kl} F p_k' \right\} (\alpha^1 \alpha^1 \alpha^2)_{imk} \cdot (\alpha^1 \alpha^2 \alpha^2)_{jlr}$$

$$= \frac{1}{32} F^2 \sum_{ikmr} p_i p_k' p_m' p_r' (\alpha^1 \alpha^1 \alpha^2)_{imk} (\alpha^1 \alpha^2 \alpha^2)_{ikr}$$

$$= 0$$

(f) Between additive and dominance by additive

$$\frac{1}{8} \sum_{\substack{ijk \\ lmn}} p_{ij} p_{kl}' p_{mn} \alpha_i^1 (\alpha^1 \alpha^1 \alpha^2)_{imk}$$

$$= \frac{1}{8} \sum_{imk} p_i p_m' p_k' \alpha_i^1 (\alpha^1 \alpha^1 \alpha^2)_{imk}$$

(g) Between additive and additive by dominance

All of the terms will be 0 as in (f).

(h) Between dominance and dominance by additive

All of the terms will be 0, similar to (f).

(i) Between dominance and additive by dominance

All of the terms will be 0, as in (h).

(j) Within additive x additive

(1) 3 different subscripts, additive in 1 parent

$$\begin{aligned}
 & \frac{1}{8} \sum_{ijkl} p_{ij} p_{kl}' (\alpha^1 \alpha^2)_{ik} (\alpha^1 \alpha^2)_{il} \\
 &= \frac{1}{8} \sum_{ikl} p_l \left\{ (1 - F) p_{kl}' p_l' + \delta_{kl} F p_k' \right\} (\alpha^1 \alpha^2)_{ik} (\alpha^1 \alpha^2)_{il} \\
 &= \frac{1}{8} F \sum_{ik} p_i p_k' \left\{ (\alpha^1 \alpha^2)_{ik} \right\}^2 \\
 &= \frac{1}{32} F \sigma_{A_1 A_2}^2
 \end{aligned}$$

(2) 3 different subscripts, additive in 2 parents

$$\frac{1}{8} \sum_{\substack{ijk \\ lrs}} p_{ij} p_{kl}' p_{rs}' (\alpha^1 \alpha^2)_{ik} (\alpha^1 \alpha^2)_{lr}$$

$$= \frac{1}{8} \sum_{ikr} p_i p'_k p'_r (\alpha^1 \alpha^2)_{ik} (\alpha^1 \alpha^2)_{ir}$$

$$= 0$$

(3) 4 different subscripts, additive in 1 parent

$$\begin{aligned} & \frac{1}{8} \sum_{ijkl} p_i p'_j p'_k p'_l (\alpha^1 \alpha^2)_{ik} (\alpha^1 \alpha^2)_{jl} \\ &= \frac{1}{8} \sum_{ijkl} \left\{ (1-F) p_i p_j + \delta_{ij} F p_i \right\} \left\{ (1-F) p'_k p'_l + \delta_{kl} F p'_k \right\} (\alpha^1 \alpha^2)_{ik} (\alpha^1 \alpha^2)_{jl} \\ &= \frac{1}{8} F^2 \sum_{ik} p_i p'_k \left\{ (\alpha^1 \alpha^2)_{ik} \right\}^2 \\ &= \frac{1}{32} F^2 \sigma^2_{A_1 A_2} \end{aligned}$$

(4) 4 different subscripts, additive in 2 parents

$$\begin{aligned} & \frac{1}{8} \sum_{\substack{ijk \\ lrs}} p_i p'_j p'_k p'_l p'_r p'_s (\alpha^1 \alpha^2)_{ik} (\alpha^1 \alpha^2)_{jr} \\ &= \frac{1}{8} \sum_{ijklr} p'_k p'_r \left\{ (1-F) p_i p_j + \delta_{ij} F p_i \right\} (\alpha^1 \alpha^2)_{ik} (\alpha^1 \alpha^2)_{jr} \\ &= \frac{1}{8} F \sum_{ikr} p_i p'_k p'_r (\alpha^1 \alpha^2)_{ik} (\alpha^1 \alpha^2)_{ir} \\ &= 0 \end{aligned}$$

(k) Between additive and additive x additive

All terms equal 0, similar to (f).

(1) Between dominance and additive x additive

All terms equal 0, similar to (h).

(m) Between dominance x additive and additive x additive

All terms equal 0, similar to (h).

(n) Within dominance by dominance

(1) 5 different subscripts

$$\begin{aligned}
 & \frac{1}{128} \sum_{\substack{ijkl \\ mnrs}} P_{ij} P_{kl}^i P_{mn} P_{rs}^i (\alpha^1 \alpha^1 \alpha^2 \alpha^2)_{imkr} (\alpha^1 \alpha^1 \alpha^2 \alpha^2)_{imks} \\
 &= \frac{1}{128} \sum_{\substack{imk \\ rs}} P_i P_k^i P_m \left\{ (1-F) P_r^i P_s^i + \delta_{rs} F P_r^i \right\} (\alpha^1 \alpha^1 \alpha^2 \alpha^2)_{imkr} (\alpha^1 \alpha^1 \alpha^2 \alpha^2)_{imks} \\
 &= \frac{1}{128} F \sum_{imkr} P_i P_k^i P_m P_r^i \left\{ (\alpha^1 \alpha^1 \alpha^2 \alpha^2)_{imkr} \right\}^2 \\
 &= \frac{1}{128} F \sigma_{D_1 D_2}^2
 \end{aligned}$$

(2) 6 different subscripts

$$\begin{aligned}
 & \frac{1}{128} \sum_{\substack{ijkl \\ mnrs}} P_{ij} P_{kl}^i P_{mn} P_{rs}^i (\alpha^1 \alpha^1 \alpha^2 \alpha^2)_{imkr} (\alpha^1 \alpha^1 \alpha^2 \alpha^2)_{inlr} \\
 &= \frac{1}{128} \sum_{\substack{ikl \\ mnrs}} P_i P_r^i \left\{ (1-F) P_k^i P_l^i + \delta_{kl} F P_k^i \right\} \left\{ (1-F) P_m P_n + \delta_{mn} F P_m \right\} (\alpha^1 \alpha^1 \alpha^2 \alpha^2)_{imkr} \\
 & \quad \cdot (\alpha^1 \alpha^1 \alpha^2 \alpha^2)_{inlr}
 \end{aligned}$$



$$= \frac{1}{128} F^2 \sum_{ikmr} p_i p'_i p_m p'_m \left\{ (\alpha^1 \alpha^1 \alpha^2 \alpha^2)_{imkr} \right\}^2$$

$$= \frac{1}{128} F^2 \sigma^2 D_1 D_2$$

(3) 7 different subscripts

$$\begin{aligned} & \frac{1}{128} \sum_{\substack{ijkl \\ mnrs}} p_{ij} p'_{kl} p_{mn} p'_{rs} (\alpha^1 \alpha^1 \alpha^2 \alpha^2)_{imkr} (\alpha^1 \alpha^1 \alpha^2 \alpha^2)_{jnls} \\ &= \frac{1}{128} \sum_{\substack{ijkl \\ mnrs}} p_m \left\{ (1-F) p_i p_j + \delta_{ij} F p_i \right\} \left\{ (1-F) p'_k p'_l + \delta_{kl} F p'_k \right\} \\ & \quad \left\{ (1-F) p'_r p'_s + \delta_{rs} F p'_r \right\} (\alpha^1 \alpha^1 \alpha^2 \alpha^2)_{imkr} (\alpha^1 \alpha^1 \alpha^2 \alpha^2)_{jnls} \\ &= \frac{1}{128} F^3 \sum_{ikmr} p_i p_m p'_k p'_r (\alpha^1 \alpha^1 \alpha^2 \alpha^2)_{imkr}^2 \\ &= \frac{1}{128} F^3 \sigma^2 D_1 D_2 \end{aligned}$$

(4) 8 different subscripts

$$\begin{aligned} & \frac{1}{128} \sum_{\substack{ijkl \\ mnrs}} p_{ij} p'_{kl} p_{mn} p'_{rs} (\alpha^1 \alpha^1 \alpha^2 \alpha^2)_{imkr} (\alpha^1 \alpha^1 \alpha^2 \alpha^2)_{jnls} \\ &= \frac{1}{128} \sum_{\substack{ijkl \\ mnrs}} \left\{ (1-F) p_i p_j + \delta_{ij} F p_i \right\} \left\{ (1-F) p'_k p'_l + \delta_{kl} F p'_k \right\} \left\{ (1-F) p_m p_n \right. \\ & \quad \left. + \delta_{mn} F p_m \right\} \left\{ (1-F) p'_r p'_s + \delta_{rs} F p'_r \right\} (\alpha^1 \alpha^1 \alpha^2 \alpha^2)_{imkr} (\alpha^1 \alpha^1 \alpha^2 \alpha^2)_{jnls} \\ &= \frac{1}{128} F^4 \sum_{ikmr} p_i p'_k p_m p'_r \left\{ (\alpha^1 \alpha^1 \alpha^2 \alpha^2)_{imkr} \right\}^2 \end{aligned}$$

$$= \frac{1}{128} F^4 \sigma_{D_1 D_2}^2$$

(c) Cross products of all terms with dominance x dominance will be 0.

A summary of these expressions is presented in Table 7. The variance between families obtained by summation of the total contribution in Table 7 is

$$\begin{aligned} & \frac{1}{2}(1+F) \sigma_A^2 + \frac{1}{4}(1+F)^2 \sigma_D^2 + \frac{1}{4}(1+F)^2 \sigma_{AA}^2 \\ & + \frac{1}{8}(1+F)^3 \sigma_{AD}^2 + \frac{1}{16}(1+F)^4 \sigma_{DD}^2. \end{aligned} \quad (22)$$

If only one locus is considered all terms except those involving  $\sigma_A^2$  and  $\sigma_D^2$  are removed and the remainder is

$$\frac{1}{2}(1+F) \sigma_A^2 + \frac{1}{4}(1+F)^2 \sigma_D^2$$

which is the same equivalent value obtained for a single locus with multiple alleles in formula (15).

#### Variance within families

The variance within families can be obtained by the same process as that used for variance between family means. Because of the tedious manipulations in this process an easier method is to obtain the variance within families as the difference between the total genotypic variance and the variance between family means.

Table 7. Contribution of square and cross product terms to variance between families

Group squares	Number	Individual contribution	Total contribution
A	8	$1/8 \sigma_A^2$	$1/2 \sigma_A^2$
D	8	$1/16 \sigma_D^2$	$1/4 \sigma_D^2$
AxD	32	$1/128 \sigma_{AD}^2$	$1/8 \sigma_{AD}^2$
AxA	16	$1/64 \sigma_{AA}^2$	$1/4 \sigma_{AA}^2$
DxD	16	$1/256 \sigma_{DD}^2$	$1/16 \sigma_{DD}^2$
Gross products			
In A	4	$1/4 F \sigma_A^2$	$1/2 F \sigma_A^2$
	24	0	0
In D	8	$1/8 F \sigma_D^2$	$1/2 F \sigma_D^2$
	4	$1/8 F^2 \sigma_D^2$	$1/4 F^2 \sigma_D^2$
	16	0	0
Bet. A and D	64	0	0
In DxA	48	$1/64 F \sigma_{AD}^2$	$3/8 F \sigma_{AD}^2$
	48	$1/64 F^2 \sigma_{AD}^2$	$3/8 F^2 \sigma_{AD}^2$
	16	$1/64 F^3 \sigma_{AD}^2$	$1/8 F^3 \sigma_{AD}^2$
	128	0	0

Table 7. (continued)

Group squares	Number	Individual contribution	Total contribution
Bet. AxD and DxA	256	0	0
Bet. A and DxA	128	0	0
Bet. A and AxD	128	0	0
Bet. D and DxA	128	0	0
Bet. D and AxD	128	0	0
In AxA	16	$1/32 F \sigma_{AA}^2$	$1/2 F \sigma_{AA}^2$
	8	$1/32 F^2 \sigma_{AA}^2$	$1/4 F^2 \sigma_{AA}^2$
	96	0	0
Bet. A and AxA	128	0	0
Bet. B and AxA	128	0	0
Bet. DxA and AxA	512	0	0
In DxD	32	$1/128 F \sigma_{DD}^2$	$1/4 F \sigma_{DD}^2$
	48	$1/128 F^2 \sigma_{DD}^2$	$3/8 F^2 \sigma_{DD}^2$
	32	$1/128 F^3 \sigma_{DD}^2$	$1/4 F^3 \sigma_{DD}^2$
	8	$1/128 F^4 \sigma_{DD}^2$	$1/16 F^4 \sigma_{DD}^2$
Products with DxD	<u>1024</u>	0	0
Total	3240		

For two loci the total genotypic variance is

$$\sigma_G^2 = \sigma_A^2 + \sigma_D^2 + \sigma_{AA}^2 + \sigma_{AD}^2 + \sigma_{DD}^2 .$$

Subtracting from this the variance between family means in equation (22), the variance within families is

$$\begin{aligned} & \left\{ 1 - \left( \frac{1+F}{2} \right) \right\} \sigma_A^2 + \left\{ 1 - \left( \frac{1+F}{2} \right)^2 \right\} \sigma_D^2 + \left\{ 1 - \left( \frac{1+F}{2} \right)^2 \right\} \sigma_{AA}^2 \\ & + \left\{ 1 - \left( \frac{1+F}{2} \right)^3 \right\} \sigma_{AD}^2 + \left\{ 1 - \left( \frac{1+F}{2} \right)^4 \right\} \sigma_{DD}^2 . \end{aligned}$$

For a single locus the  $\sigma_A^2$  and  $\sigma_D^2$  terms simplify to

$$\frac{1}{2}(1-F) \sigma_A^2 + \frac{1}{4}(1-F)(3+F) \sigma_D^2$$

which checks with equation (18) for a single locus with multiple alleles.

#### Variance between sire means

Let the arbitrary genotype of a sire be  $A_1 A_j B_k B_l$  with frequency  $P_{1j} P'_{kl}$  where

$$P_{1j} = (1-F)p_1 p_j + F\delta_{1j} p_1$$

$$P'_{kl} = (1-F)p'_k p'_l + F\delta_{kl} p'_k .$$

The dam gametic array is

$$\left\{ \sum p_r^A \right\} \left\{ \sum p_s^B \right\}$$

and the sire gametic array is

$$\left\{ \frac{1}{2} A_1 + \frac{1}{2} A_j \right\} \left\{ \frac{1}{2} B_k + \frac{1}{2} B_l \right\} .$$

The offspring array is therefore

$$\begin{aligned} & \left\{ \frac{1}{2} \sum_r p_r A_r A_1 + \frac{1}{2} \sum_r p_r A_r A_j \right\} \left\{ \frac{1}{2} \sum_s p'_s B_s B_k + \frac{1}{2} \sum_s p'_s B_s B_l \right\} \\ &= \frac{1}{4} \sum_{rs} p_r p'_s A_r A_1 B_s B_k + \frac{1}{4} \sum_{rs} p_r p'_s A_r A_1 B_s B_l \\ &+ \frac{1}{4} \sum_{rs} p_r p'_s A_r A_j B_s B_k + \frac{1}{4} \sum_{rs} p_r p'_s A_r A_j B_s B_l . \end{aligned}$$

Substitution for  $A_r A_1 B_s B_k$ ,  $A_r A_1 B_s B_l$ ,  $A_r A_j B_s B_k$ , and  $A_r A_j B_s B_l$  in terms of the genotypic values of formula (21) gives

$$\begin{aligned} & \frac{1}{4} \sum_{rs} p_r p'_s \left\{ \mu + \alpha_r^1 + \alpha_1^1 + (\alpha^1 \alpha^1)_{r1} + \alpha_s^2 + \alpha_k^2 + (\alpha^2 \alpha^2)_{sk} \right. \\ & \quad + (\alpha^1 \alpha^2)_{rs} + (\alpha^1 \alpha^2)_{rk} + (\alpha^1 \alpha^2)_{is} + (\alpha^1 \alpha^2)_{ik} \\ & \quad + (\alpha^1 \alpha^2 \alpha^2)_{rsk} + (\alpha^1 \alpha^2 \alpha^2)_{isk} + (\alpha^1 \alpha^1 \alpha^2)_{ris} + (\alpha^1 \alpha^1 \alpha^2)_{rik} \\ & \quad \left. + (\alpha^1 \alpha^1 \alpha^2 \alpha^2)_{risk} \right\} \\ & + \frac{1}{4} \sum_{rs} p_r p'_s \left\{ \mu + \alpha_r^1 + \alpha_1^1 + (\alpha^1 \alpha^1)_{r1} + \alpha_s^2 + \alpha_l^2 + (\alpha^2 \alpha^2)_{sl} \right. \\ & \quad + (\alpha^1 \alpha^2)_{rs} + (\alpha^1 \alpha^2)_{rl} + (\alpha^1 \alpha^2)_{is} + (\alpha^1 \alpha^2)_{il} \\ & \quad + (\alpha^1 \alpha^2 \alpha^2)_{rsl} + (\alpha^1 \alpha^2 \alpha^2)_{isl} + (\alpha^1 \alpha^1 \alpha^2)_{ris} + (\alpha^1 \alpha^1 \alpha^2)_{ril} \end{aligned}$$

$$\begin{aligned}
& + (\alpha^1 \alpha^1 \alpha^2 \alpha^2)_{r1s1} \} \\
& + \frac{1}{4} \sum_{rs} p_r p_s' \left\{ \mu + \alpha_r^1 + \alpha_j^1 + (\alpha^1 \alpha^1)_{rj} + \alpha_s^2 + \alpha_k^2 + (\alpha^2 \alpha^2)_{sk} \right. \\
& \quad + (\alpha^1 \alpha^2)_{rs} + (\alpha^1 \alpha^2)_{rk} + (\alpha^1 \alpha^2)_{js} + (\alpha^1 \alpha^2)_{jk} \\
& \quad + (\alpha^1 \alpha^2 \alpha^2)_{rsk} + (\alpha^1 \alpha^2 \alpha^2)_{jsk} + (\alpha^1 \alpha^1 \alpha^2)_{rjs} + (\alpha^1 \alpha^1 \alpha^2)_{rjk} \\
& \quad \left. + (\alpha^1 \alpha^1 \alpha^2 \alpha^2)_{rjsk} \right\} \\
& + \frac{1}{4} \sum_{rs} p_r p_s' \left\{ \mu + \alpha_r^1 + \alpha_j^1 + (\alpha^1 \alpha^1)_{rj} + \alpha_s^2 + \alpha_l^2 + (\alpha^2 \alpha^2)_{sl} \right. \\
& \quad + (\alpha^1 \alpha^2)_{rs} + (\alpha^1 \alpha^2)_{rl} + (\alpha^1 \alpha^2)_{js} + (\alpha^1 \alpha^2)_{jl} \\
& \quad + (\alpha^1 \alpha^2 \alpha^2)_{rsl} + (\alpha^1 \alpha^2 \alpha^2)_{jsl} + (\alpha^1 \alpha^1 \alpha^2)_{rjs} + (\alpha^1 \alpha^1 \alpha^2)_{rjl} \\
& \quad \left. + (\alpha^1 \alpha^1 \alpha^2 \alpha^2)_{rjsl} \right\} . \tag{23}
\end{aligned}$$

The summation is not performed term by term since many of the terms sum to zero. For example,

$$\sum_{rs} p_r p_s' (\alpha^1 \alpha^1)_{r1} = \sum_r p_r (\alpha^1 \alpha^1)_{r1} = 0 .$$

$$\sum_{rs} p_r p_s' (\alpha^1 \alpha^2 \alpha^2)_{rsk} = 0 ,$$

and  $\sum_{rs} p_r p_s' (\alpha^1 \alpha^1 \alpha^2 \alpha^2)_{risk} = 0 .$

Therefore, the sums all of the dominance, additive x dominance, and dominance x dominance terms equal zero. The additive and additive x additive terms are summed as

$$\sum_{rs} p_r p'_s \alpha_r^1 = \sum_r p_r \alpha_r^1 = 0,$$

$$\sum_{rs} p_r p'_s \alpha_1^1 = \alpha_1^1,$$

$$\sum_{rs} p_r p'_s (\alpha^1 \alpha^2)_{rs} = 0,$$

$$\sum_{rs} p_r p'_s (\alpha^1 \alpha^2)_{rk} = 0,$$

and  $\sum_{rs} p_r p'_s (\alpha^1 \alpha^2)_{ik} = (\alpha^1 \alpha^2)_{ik}.$

Using this procedure in equation (23), the progeny mean of  $A_1 A_j B_k B_1$  is

$$\begin{aligned} & \frac{1}{4} \mu + \frac{1}{4} \alpha_1^1 + \frac{1}{4} \alpha_k^2 + \frac{1}{4} (\alpha^1 \alpha^2)_{ik} \\ & + \frac{1}{4} \mu + \frac{1}{4} \alpha_1^1 + \frac{1}{4} \alpha_1^2 + \frac{1}{4} (\alpha^1 \alpha^2)_{i1} \\ & + \frac{1}{4} \mu + \frac{1}{4} \alpha_j^1 + \frac{1}{4} \alpha_k^2 + \frac{1}{4} (\alpha^1 \alpha^2)_{jk} \\ & + \frac{1}{4} \mu + \frac{1}{4} \alpha_j^1 + \frac{1}{4} \alpha_1^2 + \frac{1}{4} (\alpha^1 \alpha^2)_{j1} \end{aligned}$$

which simplifies to



$$\begin{aligned} \mu &+ \frac{1}{2}a_1^1 + \frac{1}{2}a_j^1 + \frac{1}{2}a_k^2 + \frac{1}{2}a_1^2 + \frac{1}{4}(a^1a^2)_{ik} \\ &+ \frac{1}{4}(a^1a^2)_{il} + \frac{1}{4}(a^1a^2)_{jk} + \frac{1}{4}(a^1a^2)_{jl} . \end{aligned} \quad (24)$$

The mean of the progeny means is  $\mu$ , because  $\sum P_{ij}P'_{kl}$  of any of the terms except  $\mu$  in equation (24) is zero, and

$$\sum P_{ij}P'_{kl}\mu = \mu .$$

The variance of sire means is therefore

$$\begin{aligned} \sum_{ijkl} P_{ij}P'_{kl} \left\{ \frac{1}{2}a_1^1 + \frac{1}{2}a_j^1 + \frac{1}{2}a_k^2 + \frac{1}{2}a_1^2 + \frac{1}{4}(a^1a^2)_{ik} \right. \\ \left. + \frac{1}{4}(a^1a^2)_{il} + \frac{1}{4}(a^1a^2)_{jk} + \frac{1}{4}(a^1a^2)_{jl} \right\}^2 . \end{aligned} \quad (25)$$

The evaluation of the squared terms is as follows:

(a) Additive

$$\frac{1}{4} \sum_{ijkl} P_{ij}P'_{kl} (a_1^1)^2 = \frac{1}{4} \sum_i P_i (a_1^1)^2 = \frac{1}{8} \sigma^2 A_1$$

(b) Additive x additive

$$\begin{aligned} &\frac{1}{16} \sum_{ijkl} P_{ij}P'_{kl} \left\{ (a^1a^2)_{ik} \right\}^2 \\ &= \frac{1}{16} \sum_{ik} P_i P'_k \left\{ (a^1a^2)_{ik} \right\}^2 \end{aligned}$$

$$= \frac{1}{64} \sigma_{AA}^2.$$

The cross product terms are also considered only by an example of each type as follows:

(a) Within additive terms

(1) both additive effects at locus 1

$$\begin{aligned} & \frac{1}{2} \sum_{ijkl} p_{ij} p_{kl}^i (\alpha_i^1) (\alpha_j^1) \\ &= \frac{1}{2} F \sum_i p_i (\alpha_i^1)^2 \\ &= \frac{1}{4} F \sigma_{A_1}^2 \end{aligned}$$

(2) effects at both loci

$$\begin{aligned} & \frac{1}{2} \sum_{ijkl} p_{ij} p_{kl}^i (\alpha_i^1) (\alpha_k^2) \\ &= \frac{1}{2} \sum_{ik} p_i p_k^i (\alpha_i^1) (\alpha_k^2) \\ &= 0 \end{aligned}$$

(b) Between additive and additive x additive

All terms are of the type

$$\frac{1}{4} \sum_{ijkl} p_{ij} p_{kl}^i (\alpha_i^1) (\alpha_i^1 \alpha_k^2)_{ik}$$

$$= 0$$

(c) Within additive x additive terms

(1) 3 alleles only

$$\begin{aligned} & \frac{1}{8} \sum_{ijkl} p_{ij} p'_{kl} (\alpha^1 \alpha^2)_{ik} (\alpha^1 \alpha^2)_{jl} \\ &= \frac{1}{8} \sum_{ik} p_i p'_k \left\{ (\alpha^1 \alpha^2)_{ik} \right\}^2 \\ &= \frac{1}{32} \sigma_{AA}^2 \end{aligned}$$

(2) 4 alleles

$$\begin{aligned} & \frac{1}{8} \sum_{ijkl} p_{ij} p'_{kl} (\alpha^1 \alpha^2)_{ik} (\alpha^1 \alpha^2)_{jl} \\ &= \frac{1}{8} \sum_{ik} p_i p'_k \left\{ (\alpha^1 \alpha^2)_{ik} \right\}^2 \\ &= \frac{1}{32} \sigma_{AA}^2 \end{aligned}$$

The terms are now added together, weighted by the number of times each appears in equation (25), to give the variance between sire means as

$$\begin{aligned} & \frac{1}{4} \sigma_{A_1}^2 + \frac{1}{4} \sigma_{A_2}^2 + \frac{1}{16} \sigma_{AA}^2 + \frac{1}{4} \sigma_{A_1}^2 + \frac{1}{4} \sigma_{A_2}^2 \\ &+ \frac{1}{8} \sigma_{AA}^2 + \frac{1}{16} \sigma_{AA}^2 \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{4} \sigma_A^2 + \frac{1}{4} F \sigma_A^2 + \frac{1}{16} \sigma_{AA}^2 + \frac{1}{8} F \sigma_{AA}^2 + \frac{1}{16} F^2 \sigma_{AA}^2 \\
&= \left( \frac{1+F}{4} \right) \sigma_A^2 + \left( \frac{1+F}{4} \right)^2 \sigma_{AA}^2 .
\end{aligned} \tag{26}$$

Under the assumption of no epistacy the additive x additive term does not appear and the remaining expression,  $\left( \frac{1+F}{4} \right) \sigma_A^2$ , is the same expression as that obtained in the case of a single locus in equation (20).

#### Variance within sires

The variance within sires is obtained by the difference between the total genotypic variance and the variance between sire means as

$$\begin{aligned}
&\sigma_A^2 + \sigma_D^2 + \sigma_{AA}^2 + \sigma_{AD}^2 + \sigma_{DD}^2 - \left( \frac{1+F}{4} \right) \sigma_A^2 - \left( \frac{1+F}{4} \right)^2 \sigma_{AA}^2 \\
&= \left\{ 1 - \left( \frac{1+F}{4} \right) \right\} \sigma_A^2 + \sigma_D^2 + \left\{ 1 - \left( \frac{1+F}{4} \right)^2 \right\} \sigma_{AA}^2 + \sigma_{AD}^2 + \sigma_{DD}^2 .
\end{aligned}$$

For a single locus this quantity simplifies to

$$\frac{1}{4}(3-F) \sigma_A^2 + \sigma_D^2$$

which checks with formula (19).

#### Extension to Arbitrary Loci

It is clear that the general expression for the covariance of full-sibs is

$$\begin{aligned}
\text{Cov(FS)} = & \left(\frac{1+F}{2}\right) \sigma_A^2 + \left(\frac{1+F}{2}\right)^2 \sigma_D^2 + \left(\frac{1+F}{2}\right)^2 \sigma_{AA}^2 \\
& + \left(\frac{1+F}{2}\right)^3 \sigma_{AD}^2 + \left(\frac{1+F}{2}\right)^4 \sigma_{DD}^2 + \left(\frac{1+F}{2}\right)^3 \sigma_{AAA}^2 \\
& + \left(\frac{1+F}{2}\right)^4 \sigma_{AAD}^2 + \left(\frac{1+F}{2}\right)^5 \sigma_{ADD}^2 + \left(\frac{1+F}{2}\right)^6 \sigma_{DDD}^2 \\
& + \text{etc.},
\end{aligned}$$

and the covariance of half-sibs is

$$\text{Cov(HS)} = \left(\frac{1+F}{4}\right) \sigma_A^2 + \left(\frac{1+F}{4}\right)^2 \sigma_{AA}^2 + \left(\frac{1+F}{4}\right)^3 \sigma_{AAA}^2 + \text{etc.}$$

A proof of these results is given by Kempthorne (38).

## CONDUCT OF EXPERIMENTS

The problem of major importance to the research worker is how to obtain estimates of the components of genotypic variance from experimental material. There are no data available in the literature which enable one to estimate all of the components indicated earlier. Estimates of  $\sigma^2_{\text{g}}$ ,  $\sigma^2_{\text{D}}$ , and  $\sigma^2_{\text{A}}$  can be obtained from six experiments conducted in Iowa in 1952 and 1953 under the assumption of no epistacy. Only one-half of the diallel table is available since reciprocal crosses were not maintained separately.

The source of material in this experiment is a 16 line synthetic variety designated Low Ear. The 16 lines which were combined into the synthetic in 1932 in the greenhouse at Arlington, Virginia were chosen because they had low ear placement on the stalk. Following is the pedigree of the synthetic, showing how the lines were combined:

$$\left[ \begin{array}{l} \{ (\text{KHLE 134-1-1-3} \times \text{Ill. LE 625-S4}) \times (\text{MWLE} \times \text{Ill. LE 630-S4}) \} \\ \{ (\text{Ohio 293-1-1-2} \times \text{Ohio 490 K}) \times (\text{Ind. WF9} \times \text{Ind. 66-24}) \} \\ \{ (\text{Ind. Tr B2} \times \text{Ind. Tr 9122}) \times (\text{Ill. A} \times \text{Ill. 90}) \} \\ \{ (\text{Ia. Cl 447 A2} \times \text{St 665}) \times (\text{US4-S} \times \text{Ia. B1356}) \} \end{array} \right] .$$

Only two of the lines, Ind. WF9 and Ill. A have been used extensively as lines, and at the present time only Ind. WF9 remains in commercial hybrids.

Following the original crosses the synthetic was grown under isolation in Iowa for at least six years and allowed to pollinate at random.

Wentworth and Remick (60) showed that for one factor equilibrium is reached in the first generation after random mating commences, regardless of the initial composition of the population. Jennings (27) showed that with two or more factors the approach to equilibrium is slowed and the rate reduced still more with linkage. Since at least six generations of random mating had occurred it was assumed that the population was in equilibrium.

In 1949 a bulk population of the synthetic was grown at Ames and plants were selfed at random. In 1950, 211 of the  $S_1$ 's were grown in progeny rows at Ankeny for evaluation of their corn borer resistance. Ten of the lines showing good resistance to borer leaf feeding were chosen for parents in a diallel series. It should be noted that no selection was practiced for plant vigor, maturity, or yield. It is felt that for combining ability these lines represent an unselected sample of  $S_1$  lines from a random mating population in equilibrium.

In 1951 from remnant seed the ten  $S_1$  lines were grown in paired rows in the breeding nursery and the 45 intercrosses were made. A large number of ears were pollinated in each pair of rows to reduce any bias from inadequate sampling and to provide enough seed for the subsequent yield tests. Seed from the crosses within each paired row were bulked to form an  $F_1$  family.

In 1952 and 1953 yield trials of the  $F_1$ 's were conducted in the North Central Section of Iowa at the locations shown in Table 8. All of the 45

Table 8. Location of diallel yield tests in 1952 and 1953

Year	District	Location	Experiment
1952	4	Storm Lake	33
	5	Clarion	34
	6	Independence	35
-----			
1953	4	Storm Lake	36
	5	Clarion	37
	6	Independence	38

$F_1$ 's were grown in three replications in each experiment. Five kernels per hill were planted in 2 x 5 hill plots. Each hill was thinned to three plants in 1952 and to four plants in 1953.

The weight of ear corn per plot was obtained at harvest and adjusted upward for missing hills by an average correction factor. This adjustment will tend to adjust the yields of plot totals to what would have been obtained with a constant number of plants,  $f$ , where  $f$  is 30 in 1952 and 40 in 1953.



## ANALYSIS OF EXPERIMENTS

## Analysis of Individual Experiments

A model that has been used to analyze the variation of a single experiment is

$$y_{ijk} = \mu + g_i + g_j + s_{ij} + r_k + e_{ijk}$$

where

$$i, j = 1, 2, \dots, p, i < j$$

$$k = 1, 2, \dots, r$$

and  $y_{ijk}$  is the yield resulting from a cross of the  $i$ -th line with the  $j$ -th line grown in the  $k$ -th replicate;  $\mu$  is a common mean for all crosses in all replicates;  $g_i$  is a measure of the average effect of the  $i$ -th line to all progeny;  $s_{ij}$  is a measure of the average effect of a cross of the  $i$ -th line and the  $j$ -th line;  $r_k$  is the average effect of all crosses grown in the  $k$ -th replicate; and  $e_{ijk}$  is the experimental error associated with the  $y_{ijk}$  observation. The component for the  $g_i$ 's is a measure of general combining ability, while the  $s_{ij}$ 's measure deviations from an additive genetic model, referred to as specific combining ability. Reciprocal crosses are assumed to be equal so only one-half of the diallel cross table is considered.

Rojas (49) considered the analysis of variance for tests of hypotheses

and for estimation of variance components. The least squares estimates of the parameters  $\mu$ ,  $\epsilon_i$ 's and  $r_k$ 's under the hypothesis of  $s_{ij}$ 's = 0 are obtained by minimizing

$$Q = \sum_{ijk} (y_{ijk} - \mu - \epsilon_i - \epsilon_j - r_k)^2.$$

The normal equations are

$$Y_{...} = nr\hat{\mu} + r(p-1) \sum_i \hat{\epsilon}_i + n \sum_k \hat{r}_k$$

$$Y_{i..} = r(p-1)\hat{\mu} + r(p-1)\hat{\epsilon}_i + r \sum_{\substack{j \\ i \neq j}} \hat{\epsilon}_j + (p-1) \sum_k \hat{r}_k$$

$$Y_{..k} = nr\hat{\mu} + (p-1) \sum_i \hat{\epsilon}_i + nr\hat{r}_k$$

which lead to a solution

$$\hat{\mu} = \frac{Y_{...}}{nr}$$

$$\hat{\epsilon}_i = \frac{1}{r(p-2)} \left\{ Y_{i..} - \frac{2Y_{...}}{p} \right\}$$

$$\hat{r}_k = \frac{Y_{..k}}{n} - \frac{Y_{...}}{rn}$$

The sums of squares obtained by the products of the estimates and the right hand sides of the normal equations are as follows:

$$n = \frac{p(p-1)}{2}$$

$$C = \frac{Y^2_{...}}{rn}$$

$$R = \sum_k \frac{y_{..k}^2}{n} - C$$

$$V = \sum_{ij} \frac{y_{ij.}^2}{r} - C$$

$$G = \frac{1}{r(p-2)} \sum_i y_{i..}^2 - \frac{2(p-1)C}{p-2}$$

$$T = \sum_{ijk} y_{ijk}^2 - C.$$

The procedure for getting the separate components of the total sum of squares was given by Kempthorne (33, p. 113) for a similar but more complex example.

The lines are to be chosen at random from some population so the  $g_i$ 's are considered as a population of general combining ability values and the  $s_{ij}$ 's as a population of specific combining ability values.

The expected mean squares are obtained by finding the expected values of the sums of squares in terms of the original model and dividing by the degrees of freedom. The values obtained by Rojas are given in Table 9.

In the estimation of variance components from the expected mean squares, it is assumed that the  $g_i$ 's are NID  $(0; \sigma_g^2)$ , the  $s_{ij}$ 's are NID  $(0; \sigma_s^2)$ , the  $e_{ijk}$ 's are NID  $(0; \sigma^2)$ , and  $\mu$  and the  $r_k$ 's are constants. The assumption of normality is required only for purposes of tests of significance.

Table 9. Analysis of variance and expected mean squares for  $\sigma_g^2$  and  $\sigma_s^2$ 

Source	d.f.	S.S.	M.S.	E.M.S.
Replicates	r-1	R	R'	
<hr/>				
$\bar{g}_i$ 's	p-1	G	G'	$\sigma^2 + r\sigma_s^2 + r(p-2)\sigma_g^2$
$s_{ij}$ 's	p(p-3)/2	S = V - G	S'	$\sigma^2 + r\sigma_s^2$
<hr/>				
Grosses	n-1	V	V'	
Error	(r-1)(n-1)	E = T - (V + R)	E'	$\sigma^2$
Total	rn-1	T		

The expected mean squares given by Rojas were in terms of variances of general and specific combining ability. They will now be derived in terms of variances and covariances of relatives, in respect to a random mating population by again taking the expectation of the sums of squares in Table 9.

Consider a diallel table where a plot is the progeny of a mating of two parents with a known amount of inbreeding. Variances and covariances of plot totals will be obtained for a constant number,  $f$ , plants per plot.

All of the individuals in a plot have the same sire and dam so the plot total contains  $f$  plants which are full-sibs. The expected value of a plot total is  $f u_g$  where  $u_g$  is the mean of the genotypic portion of the

individuals. Then

$$E(\text{plot total})^2 = f^2 \sigma_G^2 + f \sigma_G^2 + f(f-1) \text{Cov}(\text{FS})$$

where  $\sigma_G^2$  is the genotypic variance and  $\text{Cov}(\text{FS})$  is the covariance of full-sibs. The variance of a plot total is obtained by the usual method as

$$\begin{aligned} V(\text{plot total}) &= E(\text{plot total})^2 - \{E(\text{plot total})\}^2 \\ &= f \sigma_G^2 + f(f-1) \text{Cov}(\text{FS}) = \sigma_1^2. \end{aligned} \quad (27)$$

Next consider the covariance of two plots, both with the same sire and dam, but in different replicates. If the parents are not homozygous the individuals in one replicate are not a true duplication of the genotypic value of individuals in the same entry in another replicate. In fact, the individuals are full-sibs. With  $f$  plants per plot, the covariance between two plots with the same parents, in different replicates, is

$$f^2 \text{Cov}(\text{FS}) = \sigma_2^2. \quad (28)$$

The covariance between plots with one parent in common will be a function of the half-sib covariance. With  $f$  individuals per plot this covariance will be

$$f^2 \text{Cov}(\text{HS}) = \sigma_3^2 \quad (29)$$

where  $\text{Cov}(\text{HS})$  is the covariance of half-sibs.

The expected mean squares will be derived in terms of  $\sigma_1^2$ ,  $\sigma_2^2$  and  $\sigma_3^2$ . The mean is coded as zero so it will not enter into the derivation. Replicates are orthogonal to crosses so their effects can be ignored in most of the computations below. Terms involving both the mean and replication will cancel out if carried along.

The  $Y_{1..}$  term, a sire total, contains  $r(p - 1)$  plots, and  $Y_{1..}^2$  will involve  $r(p - 1)\sigma_1^2$  terms. Each of the  $(p - 1)$  plot totals is a full-sib to  $r(r - 1)$  plot totals in other replicates. Each of  $(p - 1)$  plot totals in one replication is a half-sib with  $(p - 2)$  plots in  $r$  replicates, over the total of  $r$  replicates. Then

$$E(Y_{1..}^2) = r(p - 1)\sigma_1^2 + r(r - 1)(p - 1)\sigma_2^2 + r^2(p - 1)(p - 2)\sigma_3^2.$$

The grand total,  $Y_{...}$ , contains a total of  $rn$  plots. The square of this term will contain  $rn$  terms of  $\sigma_1^2$ . The full-sib comparisons will be  $p/2$  times those for  $Y_{1..}$  since all of the dams are also included. Each of  $n$  plot totals in one replicate is a half-sib of  $2(p - 2)$  plots in  $r$  replicates, over  $r$  replicates, and

$$E(Y_{...}^2) = rn\sigma_1^2 + r(r - 1)n\sigma_2^2 + 2r^2(p - 2)n\sigma_3^2.$$

Each  $Y_{1j.}$  term contains  $r$  plots, so  $Y_{1j.}^2$  has  $r$  terms of  $\sigma_1^2$ . The full-sib covariances will enter between the replicates  $r(r - 1)$  times, and there are no half-sib covariances. Then

$$E(Y_{1j.}^2) = r\sigma_1^2 + r(r - 1)\sigma_2^2.$$

A total of  $n$  plots are contained in each  $Y_{..k}$  term. In  $Y_{..k}^2$  there will be  $n$  terms of  $\sigma_1^2$ , and no terms in  $\sigma_2^2$  because there are no comparisons between replicates. Each of  $n$  plot totals will involve half-sib covariances with  $2(p-2)$  plots resulting in

$$Y_{..k}^2 = n\sigma_1^2 + 2n(p-2)\sigma_3^2 + nr_k^2.$$

The expected values of the sums of squares will now be obtained from the sums of squares in Table 9.

$$E(C) = \frac{1}{rn} E(Y_{..k}^2)$$

$$= \sigma_1^2 + (r-1)\sigma_2^2 + 2r(p-2)\sigma_3^2$$

$$E(R) = \frac{1}{n} E\left(\sum_k Y_{..k}^2\right) - E(C)$$

$$= (r-1)\sigma_1^2 - (r-1)\sigma_2^2 + \frac{n}{r-1} \sum r_k^2$$

$$E(G) = \frac{1}{r(p-2)} E\left(\sum_i Y_{i..}^2\right) - \frac{2(p-1)}{p-2} E(C)$$

$$= (p-1)\sigma_1^2 + (p-1)(r-1)\sigma_2^2 + r(p-1)(p-4)\sigma_3^2$$

$$E(S) = \frac{1}{r} E\left(\sum_{ij} Y_{ij.}^2\right) - E(C) - E(G)$$

$$= \frac{p(p-3)}{2} \sigma_1^2 + \frac{(r-1)p(p-3)}{2} \sigma_2^2 - rp(p-3)\sigma_3^2.$$

The total sum of squares contains  $rn$  plots each with an expected

value of  $\sigma_1^2$  when squared. Thus

$$\begin{aligned} E(T) &= E\left(\sum_{ijk} y_{ijk}^2\right) - E(C) \\ &= (rn - 1)\sigma_1^2 - (r - 1)\sigma_2^2 - 2r(p - 2)\sigma_3^2. \end{aligned}$$

The expected sum of squares for error obtained by difference is

$$(r - 1)(n - 1)\sigma_1^2 - (r - 1)(n - 1)\sigma_2^2 - \sigma_3^2.$$

The expected sums of squares in terms of  $\sigma_1^2$ ,  $\sigma_2^2$ , and  $\sigma_3^2$  are presented in Table 10.

Dividing the expected sums of squares by the degrees of freedom gives the expected mean squares in Table 11 after slight rearrangement. The replicates source of variation also has a term,  $\frac{n}{r-1} \sum_k r_k^2$ . Each component of variation contains a term for plot error ( $\sigma_e^2$ ).

The similarity between Tables 9 and 11 is very evident. The exact relationship is

$$\begin{aligned} \sigma^2 &= \sigma_e^2 + \sigma_1^2 - \sigma_2^2 \\ \sigma_s^2 &= \sigma_2^2 - 2\sigma_3^2 \\ \sigma_g^2 &= \sigma_3^2. \end{aligned}$$

After substituting for  $\sigma_1^2$ ,  $\sigma_2^2$ , and  $\sigma_3^2$  from equations (27), (28), and (29), the equality is



Table 10. Expected sums of squares for components of genotypic variance in a single experiment

Source	d.f.	E.S.S.		
		$\sigma_1^2$	$\sigma_2^2$	$\sigma_3^2$
Replicates	r-1	r-1	-(r-1)	0
$\bar{g}_i$ 's	p-1	p-1	(p-1)(r-1)	r(p-1)(p-r)
$\bar{g}_{ij}$ 's	p(p-3)/2	p(p-3)/2	(r-1)p(p-3)/2	-rp(p-3)
Error	(r-1)(n-1)	(r-1)(n-1)	-(r-1)(n-1)	0
Total	rn-1	rn-1	-(r-1)	-2r(p-2)

Table 11. Expected mean squares for components of genotypic variance in a single experiment

Source	d.f.	E.M.S.		
		$\sigma_1^2 - \sigma_2^2$	$\sigma_2^2 - 2\sigma_3^2$	$\sigma_3^2$
Replicates	r-1	1	0	0
$\bar{g}_i$ 's	p-1	1	r	r(p-2)
$\bar{g}_{ij}$ 's	p(p-3)/2	1	r	0
Error	(r-1)(n-1)	1	0	0
Total	rn-1			

$$\sigma^2 = \sigma_e^2 + r \left\{ \sigma_g^2 - \text{Cov}(FS) \right\}$$

$$\sigma_s^2 = r^2 \left\{ \text{Cov}(FS) - 2\text{Cov}(HS) \right\} \quad (30)$$

$$\sigma_g^2 = r^2 \text{Cov}(HS). \quad (31)$$

Estimates of these variance components are obtained by equating the expected mean square to the observed mean square and solving for the desired component. The estimates are obtained as

$$\hat{\sigma}_g^2 = \frac{Q' - S'}{r(p-2)}$$

$$\hat{\sigma}_s^2 = \frac{S' - E'}{r}$$

$$\hat{\sigma}_e^2 = E'.$$

The  $\text{Cov}(FS)$  and  $\text{Cov}(HS)$  can be obtained from these estimates as

$$\hat{\text{Cov}}(HS) = \frac{\hat{\sigma}_s^2}{r^2}$$

$$\hat{\text{Cov}}(FS) = \frac{\hat{\sigma}_s^2 + 2\hat{\sigma}_e^2}{r^2}.$$

From the theoretical derivation of covariances in the previous sections of this thesis, two of the variances are utilized here. The variance between family means, equation (22), is  $\text{Cov}(FS)$ , and the variance between sire means, equation (26), is  $\text{Cov}(HS)$ .

Making the assumption of no epistacy these variances are

$$\text{Cov}(\text{HS}) = \left(\frac{1+F}{4}\right)\sigma_A^2$$

and

$$\text{Cov}(\text{FS}) = \left(\frac{1+F}{2}\right)\sigma_A^2 + \left(\frac{1+F}{2}\right)^2\sigma_D^2.$$

Estimates of  $\sigma_A^2$ ,  $\sigma_D^2$  and  $\sigma_G^2$  are obtained as

$$\hat{\sigma}_A^2 = \frac{4 \hat{\text{Cov}}(\text{HS})}{(1+F)}$$

$$\hat{\sigma}_D^2 = \frac{4 \hat{\text{Cov}}(\text{FS}) - 2 \hat{\text{Cov}}(\text{HS})}{(1+F)^2}$$

and

$$\hat{\sigma}_G^2 = \hat{\sigma}_A^2 + \hat{\sigma}_D^2.$$

An estimate of  $\sigma_e^2$  may be obtained by the equality

$$\sigma_e^2 = \sigma^2 + \sigma_2^2 - \sigma_1^2.$$

In the present experiment there are two sources of environmental variance. These are (1) between plants in plots,  $\sigma_w^2$ , and (2) between plots,  $\sigma_p^2$ . Thus the error for the total of a plot of  $f$  plants is

$$\sigma_e^2 = f^2\sigma_p^2 + f\sigma_w^2.$$

In the present experiment it is impossible to estimate  $\sigma_p^2$  and  $\sigma_w^2$  separately because individual plant data were not recorded. If measurements on individual plants were available, two heritability

estimates could be obtained. These are

$$H_1 = \frac{\sigma_A^2}{\sigma_G^2 + \sigma_W^2}$$

and

$$H_2 = \frac{\sigma_A^2}{\sigma_G^2 + \sigma_W^2 + \sigma_P^2}.$$

$H_1$  is an estimate of heritability where individuals are planted within plots, and  $H_2$  is the estimate where the individuals are in different plots.

The values of  $\sigma_G^2$  and  $\sigma_E^2$  may be expressed in terms of additive, dominance, and epistatic variances to study the type of epistatic deviation which may bias the estimates. The estimates obtained from equations (30) and (31) are

$$\sigma_G^2 = f^2 \left\{ \left( \frac{1+F}{4} \right) \sigma_A^2 + \left( \frac{1+F}{4} \right)^2 \sigma_{AA}^2 \right\}$$

and

$$\sigma_E^2 = f^2 \left\{ \left( \frac{1+F}{2} \right)^2 \sigma_D^2 + \frac{(1+F)^2}{8} \sigma_{AA}^2 + \left( \frac{1+F}{2} \right)^3 \sigma_{AD}^2 + \left( \frac{1+F}{2} \right)^4 \sigma_{DD}^2 \right\}.$$

#### Analysis of Experiments Combined over Locations and Years

Of major interest is the estimation of the components of genotypic

variance over several locations and years, and the components of variance due to possible interactions of genotypes with each.

An extension of the model for a single experiment to include locations and years is

$$\begin{aligned}
 Y_{ijkltq} = & \mu + \epsilon_i + \epsilon_j + s_{ij} + d_k + \gamma_t + (gd)_{ik} + (gd)_{jk} \\
 & + (sd)_{ijk} + (gy)_{it} + (gy)_{jt} + (sy)_{ijt} + (dy)_{kt} \\
 & + (gdy)_{ikt} + (gdy)_{jkt} + (sdy)_{ijkt} + r_{ktq} + e_{ijkltq}
 \end{aligned}$$

where  $Y_{ijkltq}$  is the yield of a cross of the  $i$ -th and  $j$ -th lines in the  $q$ -th replicate in the  $k$ -th location in the  $t$ -th year;  $\mu$  is a common mean for all crosses in  $r$  replicates,  $b$  locations, and  $c$  years;  $\epsilon_i$  is a measure of the average effect of the  $i$ -th line to all progeny;  $s_{ij}$  is a measure of the average effect of a cross of the  $i$ -th line to the  $j$ -th line;  $d_k$  is the average effect of the  $k$ -th location;  $\gamma_t$  is the average effect of the  $t$ -th year;  $(gd)_{ik}$  is the average interaction of the  $i$ -th line with the  $k$ -th location; and the other interactions have the appropriate meaning designated by the subscripts. In all cases the averages are over the populations of other factors.

For purposes of this study all parameters in the model except  $\mu$  and  $r_{ktq}$  are considered random with expectation zero and zero correlation.

The subscripts take the range of values

$$i, j = 1, 2, \dots, p, \quad i < j$$

$$k = 1, 2, \dots, b$$

$$t = 1, 2, \dots, c$$

$$i = 1, 2, \dots, r$$

Rojas (49) presented the normal equations and least squares estimates of the above model and obtained the analysis of variance of Table 12. The sums of squares are obtained as follows:

$$C = \frac{Y^2}{nrbc}$$

$$Y = \sum_t \frac{Y^2_{\dots t.}}{nr b} - C$$

$$D = \sum_k \frac{Y^2_{\dots k.}}{nr c} - C$$

$$V = \sum_{ij} \frac{Y^2_{ij \dots}}{rbc} - C$$

$$VY = \sum_{ijt} \frac{Y^2_{ij.t.}}{br} - C - V - Y$$

$$VD = \sum_{ijk} \frac{Y^2_{ijk.}}{nr} - C - V - D$$

$$DY = \sum_{kt} \frac{Y^2_{\dots kt.}}{nr} - C - D - Y$$

$$VDY = \sum_{ijkt} \frac{Y^2_{ijkt.}}{r} - C - D - V - Y - VY - VD - DY$$

Table 12. Analysis of variance of combined data

Source	d.f.	S.S.	M.S.
$\bar{g}_1$ 's	p-1	G	G'
$s_{ij}$ 's	p(p-3)/2	S = V - G	S'
Crosses	(n-1)	V	V'
Locations	(b-1)	D	D'
Years	(c-1)	Y	Y'
Locations x Years	(b-1)(c-1)	DY	DY'
$(gd)_{ik}$ 's	(p-1)(b-1)	GD	GD'
$(sd)_{ijk}$ 's	(b-1)p(p-3)/2	SD = VD - GD	SD'
Crosses x Locations	(n-1)(b-1)	VD	VD'
$(gy)_{it}$ 's	(c-1)(p-1)	GY	GY'
$(sy)_{ijt}$ 's	(c-1)p(p-3)/2	SY = VY - GY	SY'
Crosses x Years	(n-1)(c-1)	VY	VY'
$(gdy)_{ikt}$ 's	(b-1)(c-1)(p-1)	GDY	GDY'
$(sdy)_{ijkt}$ 's	(b-1)(c-1)p(p-3)/2	SDY = VDY - GDY	SDY'
Crosses x Loc. x Years	(n-1)(b-1)(c-1)	VDY	VDY'
Error	(n-1)(r-1)bc	E	E'

$$G = \sum_i \frac{Y_{1....}^2}{rbc(p-2)} - \frac{2(p-1)C}{p-2}$$

$$GD = \sum_{ik} \frac{Y_{1.k..}^2}{rc(p-2)} - \frac{2(p-1)C}{p-2} - G - \frac{2(p-1)D}{p-2}$$

$$GY = \sum_{ikt} \frac{Y^2_{1..t.}}{r(p-2)} - \frac{2(p-1)C}{p-2} - G - \frac{2(p-1)Y}{p-2}$$

$$GDY = \sum_{ikt} \frac{Y^2_{1.kt.}}{r(p-2)} - \frac{2(p-1)C}{p-2} - G - \frac{2(p-1)D}{p-2} - \frac{2(p-1)Y}{p-2}$$

$$- \frac{2(p-1)DY}{p-2} - GD - GY .$$

The values for S and its interactions are obtained by difference, and E is obtained by pooling the error mean squares of the individual experiments.

The expected mean squares are presented in Table 13.

The expected mean squares in Table 11 can be extended over locations and years by observation from Table 13 because crosses and hence comparisons of the  $g_i$ 's or  $s_{ij}$ 's are orthogonal to locations and years. The expected mean squares in terms of genotypic components are given in Table 14.

The values for  $\sigma_1^2$ ,  $\sigma_2^2$ , and  $\sigma_3^2$  have the same meanings as in the individual experiments. The addition of  $d$  or  $y$  to these terms indicates the corresponding interactions. For example,  $\sigma_{2dy}^2$  is the three-way interaction of locations and years with a pseudo-factor  $x$ , say, which gives rise to the variance  $\sigma_2^2$ .

Because of the relationship between the expected mean squares in Table 13 and Table 14 it is not necessary to obtain the estimates of  $\sigma_2^2$ ,  $\sigma_3^2$ , and their interaction components. The estimate of  $\text{Cov}(HS)$ ,



Table 13. Components of variance of general and specific combining ability over locations and years

Source	E.M.S.
$\xi_1$ 's	$\sigma^2 + r\sigma_{sdy}^2 + b\sigma_{sy}^2 + c\sigma_{sd}^2 + b\sigma_s^2 + (p-2)r\sigma_{gdy}^2 + (p-2)b\sigma_{gy}^2 + (p-2)c\sigma_{gd}^2 + (p-2)b\sigma_g^2$
$s_{ij}$ 's	$\sigma^2 + r\sigma_{sdy}^2 + b\sigma_{sy}^2 + c\sigma_{sd}^2 + b\sigma_s^2$
$d_k$ 's	$\sigma^2 + r\sigma_{sdy}^2 + 2(p-1)r\sigma_{gdy}^2 + n\sigma_{dy}^2 + 2(p-1)c\sigma_{gd}^2 + c\sigma_{sd}^2 + n\sigma_d^2$
$y_t$ 's	$\sigma^2 + r\sigma_{sdy}^2 + 2(p-1)r\sigma_{gdy}^2 + n\sigma_{dy}^2 + b\sigma_{sy}^2 + 2(p-1)b\sigma_{gy}^2 + n\sigma_y^2$
$(dy)_{kt}$ 's	$\sigma^2 + r\sigma_{sdy}^2 + 2(p-1)r\sigma_{gdy}^2 + n\sigma_{dy}^2$
$(gd)_{ik}$ 's	$\sigma^2 + r\sigma_{sdy}^2 + c\sigma_{sd}^2 + (p-2)r\sigma_{gdy}^2 + (p-2)c\sigma_{gd}^2$
$(sd)_{ijk}$ 's	$\sigma^2 + r\sigma_{sdy}^2 + c\sigma_{sd}^2$
$(gy)_{it}$ 's	$\sigma^2 + r\sigma_{sdy}^2 + b\sigma_{sy}^2 + (p-2)r\sigma_{gdy}^2 + (p-2)b\sigma_{gy}^2$
$(sy)_{ijt}$ 's	$\sigma^2 + r\sigma_{sdy}^2 + b\sigma_{sy}^2$
$(gdy)_{ikt}$ 's	$\sigma^2 + r\sigma_{sdy}^2 + (p-2)r\sigma_{gdy}^2$
$(sdy)_{ijkt}$ 's	$\sigma^2 + r\sigma_{sdy}^2$
Error	$\sigma^2$

Table 14. Expected mean squares for components of genotypic variance over locations and years

Source	E.M.S. <sup>a</sup>
$\epsilon_1$ 's	$\sigma_4^2 + r\sigma_{5dy}^2 + br\sigma_{5y}^2 + cr\sigma_{5d}^2 + bcr\sigma_5^2 + r(p-2)\sigma_{3dy}^2$ $+ br(p-2)\sigma_{3y}^2 + cr(p-2)\sigma_{3d}^2 + bcr(p-2)\sigma_3^2$
$\epsilon_{1j}$ 's	$\sigma_4^2 + r\sigma_{5dy}^2 + br\sigma_{5y}^2 + cr\sigma_{5d}^2 + bcr\sigma_5^2$
$(gd)_{1k}$ 's	$\sigma_4^2 + r\sigma_{5dy}^2 + cr\sigma_{5d}^2 + r(p-2)\sigma_{3dy}^2 + cr(p-2)\sigma_{3d}^2$
$(sd)_{1jk}$ 's	$\sigma_4^2 + r\sigma_{5dy}^2 + cr\sigma_{5d}^2$
$(gy)_{1t}$ 's	$\sigma_4^2 + r\sigma_{5dy}^2 + br\sigma_{5y}^2 + r(p-2)\sigma_{3dy}^2 + br(p-2)\sigma_{3y}^2$
$(sy)_{1jt}$ 's	$\sigma_4^2 + r\sigma_{5dy}^2 + br\sigma_{5y}^2$
$(gdy)_{1kt}$ 's	$\sigma_4^2 + r\sigma_{5dy}^2 + r(p-2)\sigma_{3dy}^2$
$(sdy)_{1jkt}$ 's	$\sigma_4^2 + r\sigma_{5dy}^2$
Error	$\sigma_4^2$

<sup>a</sup>Expected mean squares coded as

$$\sigma_4^2 = \sigma_e^2 + \sigma_1^2 - \sigma_2^2$$

$$\sigma_{5dy}^2 = \sigma_{2dy}^2 - 2\sigma_{3dy}^2$$

$$\sigma_{5y}^2 = \sigma_{2y}^2 - 2\sigma_{3y}^2$$

$$\sigma_{5d}^2 = \sigma_{2d}^2 - 2\sigma_{3d}^2$$

$$\sigma_5^2 = \sigma_2^2 - 2\sigma_3^2$$

$\text{Cov}(\text{FS})$ , and all interaction terms can be obtained from the following relationships:

$$\sigma_g^2 = \sigma_3^2 = f^2 \text{Cov}(\text{HS})$$

$$\sigma_s^2 = \sigma_5^2 = f^2 \{ \text{Cov}(\text{FS}) - 2\text{Cov}(\text{HS}) \}$$

$$\sigma_{gd}^2 = \sigma_{3d}^2 = f^2 \text{Cov}(\text{HS})_d$$

$$\sigma_{sd}^2 = \sigma_{5d}^2 = f^2 \{ \text{Cov}(\text{FS})_d - 2\text{Cov}(\text{HS})_d \}$$

$$\sigma_{gy}^2 = \sigma_{3y}^2 = f^2 \text{Cov}(\text{HS})_y$$

$$\sigma_{sy}^2 = \sigma_{5y}^2 = f^2 \{ \text{Cov}(\text{FS})_y - 2\text{Cov}(\text{HS})_y \}$$

$$\sigma_{gdy}^2 = \sigma_{3dy}^2 = f^2 \text{Cov}(\text{HS})_{dy}$$

$$\sigma_{sdy}^2 = \sigma_{5dy}^2 = f^2 \{ \text{Cov}(\text{FS})_{dy} - 2\text{Cov}(\text{HS})_{dy} \}$$

$$\sigma^2 = \sigma_4^2 = \sigma_e^2 + f \{ \sigma_g^2 - \text{Cov}(\text{FS}) \} .$$

The term  $\text{Cov}(\text{HS})_d$ , for example, is the component for the interaction of the likeness of half-sibs with locations. Estimates of the components of variance are obtained by equating the expected mean squares to the observed mean squares and solving for the desired components.

Assuming no epistacy, estimates of  $\sigma_A^2$ ,  $\sigma_D^2$ , and the interaction of additive effects and dominance deviations with locations and years can be obtained by the following equations:

$$\hat{\sigma}_A^2 = \frac{4\hat{\text{Cov}}(\text{HS})}{(1+F)}$$

$$\hat{\sigma}_D^2 = \frac{4 \hat{\text{Cov}}(\text{FS}) - 2\hat{\text{Cov}}(\text{HS})}{(1+F)^2}$$

$$\hat{\sigma}_{Ay}^2 = \frac{4\hat{\text{Cov}}(\text{HS})_y}{(1+F)}$$

$$\hat{\sigma}_{Dy}^2 = \frac{4 \hat{\text{Cov}}(\text{FS})_y - 2\hat{\text{Cov}}(\text{HS})_y}{(1+F)^2}$$

$$\hat{\sigma}_{Ad}^2 = \frac{4 \hat{\text{Cov}}(\text{HS})_d}{(1+F)}$$

$$\hat{\sigma}_{Dd}^2 = \frac{4 \hat{\text{Cov}}(\text{FS})_d - 2\hat{\text{Cov}}(\text{HS})_d}{(1+F)^2}$$

$$\hat{\sigma}_{A dy}^2 = \frac{4\hat{\text{Cov}}(\text{HS})_{dy}}{(1+F)}$$

$$\hat{\sigma}_{D dy}^2 = \frac{4 \hat{\text{Cov}}(\text{FS})_{dy} - 2\hat{\text{Cov}}(\text{HS})_{dy}}{(1+F)^2}$$

$$\hat{\sigma}_G^2 = \hat{\sigma}_A^2 + \hat{\sigma}_D^2$$

$$\hat{\sigma}_{Gy}^2 = \hat{\sigma}_{Ay}^2 + \hat{\sigma}_{Dy}^2$$

$$\hat{\sigma}_{Gd}^2 = \hat{\sigma}_{Ad}^2 + \hat{\sigma}_{Dd}^2$$

$$\hat{\sigma}_{G dy}^2 = \hat{\sigma}_{A dy}^2 + \hat{\sigma}_{D dy}^2$$

## RESULTS

## Individual Experiments

The six experiments were first analyzed by the method illustrated in Table 9, without subdivision of the crosses sum of squares. These results are presented in Table 15.

A difference between crosses was significant at the 1 percent level in all six experiments. The errors appeared to be quite consistent over all experiments.

Estimates of  $\sigma_s^2$  and  $\sigma_g^2$  obtained by the analysis of variance of Table 9 are presented in Table 16 together with the estimates of  $\sigma_s^2 + \sigma_1^2$ ,  $\sigma_2^2$ , and  $\sigma_3^2$  obtained by the expectations in Table 11. The mean plot yield in pounds is also given in Table 16. Significant deviations from zero, tested by an approximate F-test due to Satterthwaite (51), are indicated.

Evidence of deviation from zero was given for all values of  $\sigma_g^2$  except district 6 in 1952 and for all values of  $\sigma_s^2$  except district 5 in 1952. In four of the six experiments  $\sigma_s^2$  was numerically larger than  $\sigma_g^2$ . In all experiments the estimate of  $\sigma_2^2$  deviated significantly from zero.

Estimates of  $\text{Cov}(\text{HS})$ ,  $\text{Cov}(\text{FS})$ ,  $\sigma_A^2$ ,  $\sigma_D^2$ , and  $\sigma_G^2$  are presented in

Table 15. Analysis of variance of crosses in six diallel cross experiments in 1952 and 1953

Year	Source	d.f.	District		
			M.S.		
			4	5	6
1952	Replicates	2	8.04	2.30	30.43
	Crosses	44	7.08 <sup>a</sup>	5.82 <sup>a</sup>	7.36 <sup>a</sup>
	Error	88	2.69	3.06	3.63
1953	Replicates	2	27.41	4.65	119.14
	Crosses	44	7.21 <sup>a</sup>	7.18 <sup>a</sup>	9.86 <sup>a</sup>
	Error	88	2.03	1.43	2.36

<sup>a</sup>Denotes significance at 1 percent level.

Table 16. Means and components of variance of six diallel cross experiments in 1952 and 1953

Year	District	Mean	$\sigma^2_g$ or $\sigma^2_3$	$\sigma^2_s$	$\sigma^2_2$	$\sigma^2_e + \sigma^2_1$
1952	4	16.42	.48 <sup>a</sup>	.68 <sup>b</sup>	1.64 <sup>a</sup>	4.33
	5	14.58	.40 <sup>a</sup>	.27 <sup>b</sup>	1.07 <sup>a</sup>	4.12
	6	17.01	.23	.86 <sup>b</sup>	1.33 <sup>a</sup>	4.96
1953	4	15.39	.50 <sup>a</sup>	.91 <sup>a</sup>	1.91 <sup>a</sup>	3.94
	5	12.45	.77 <sup>a</sup>	.66 <sup>a</sup>	2.19 <sup>a</sup>	3.63
	6	11.65	.68 <sup>a</sup>	1.38 <sup>a</sup>	2.75 <sup>a</sup>	5.10

<sup>a</sup>Denotes significance at 1 percent level.<sup>b</sup>Denotes significance at 5 percent level.

Table 17. The experiments in 1952 were thinned to 30 plants per plot and in 1953 to 40 plants per plot. This changed the value of  $f$  between the two years in the calculation of  $\text{Cov}(\text{HS})$  and the  $\text{Cov}(\text{FS})$ . The value of  $F = \frac{1}{2}$  was used in obtaining  $\sigma_A^2$  and  $\sigma_D^2$  because the parents in the

Table 17. Estimates of components of genotypic variance for individual experiments

Year	District	$\text{Cov}(\text{HS})^a$	$\text{Cov}(\text{FS})^a$	$\sigma_A^2$	$\sigma_D^2$	$\frac{\sigma_D^2}{\sigma_A^2}$	$\frac{\sigma_A^2}{\sigma_G^2}$
1952	4	.53	1.82	1.41	1.35	0.96	.51
	5	.44	1.18	1.18	0.53	0.44	.69
	6	.26	1.47	0.68	1.71	2.49	.29
-----							
1953	4	.31	1.19	0.83	1.01	1.22	.45
	5	.48	1.37	1.28	0.73	0.57	.63
	6	.43	1.72	1.14	1.54	1.35	.42
Average				1.09	1.14		.50

<sup>a</sup>Coded by 3 decimals to the right

diallel table were first generation selfs of individuals in a random mating population.

In three of the experiments the estimate of  $\sigma_A^2$  was numerically larger than  $\sigma_D^2$ , and in the other three experiments  $\sigma_D^2$  was larger. The average values of  $\sigma_A^2$  and  $\sigma_D^2$  over the six experiments were almost identical. As an average of the six experiments  $\sigma_A^2$  accounted for 50

percent of the total genotypic variance, of course under the assumption of negligible epistasy.

#### Results Combined over Locations and Years

The combined analysis of variance is presented in Table 15. The estimate of error mean square was obtained by pooling the error mean square of the individual experiments.

The estimate of  $\sigma_y^2$  was extremely large in relation to the other estimates but the approximate F value with 1 degree of freedom was not significant at the 5 percent level. There is no evidence of a difference between locations, however the year x location interaction component was significant at the 1 percent level.

The estimate of  $\sigma_g^2$  was almost negligible in the combined analysis while the estimate of  $\sigma_s^2$  deviated significantly from zero. There was no evidence of an interaction of  $\sigma_s^2$  with either years or locations. The interaction of  $\sigma_g^2$  with years and the three factor interaction were significant but there was no indication of an interaction of  $\sigma_g^2$  with locations.

To obtain estimates of Cov(FS), Cov(HS), and their interactions with locations and years an average number of plants per plot were used for all of the components. Since one-half of the experiments were thinned to 30 plants per plot and the other one-half to 40 plants, a mean of 35 plants per plot was used for the value of f. The estimates of Cov(FS)



Table 18. Combined analysis of variance

Source	d.f.	M.S.	Component	Estimate
$y_t$ 's	1	1637.26	$\sigma_y^2$	9.13
$d_k$ 's	2	396.05	$\sigma_d^2$	0.20
$(dy)_{kt}$ 's	2	343.38	$\sigma_{dy}^2$	2.47 <sup>a</sup>
$g_i$ 's	9	48.38	$\sigma_g^2$	0.04
$s_{ij}$ 's	35	13.96	$\sigma_s^2$	0.58 <sup>a</sup>
$(gy)_{it}$ 's	9	32.90	$\sigma_{gy}^2$	0.38 <sup>a</sup>
$(sy)_{ijt}$ 's	35	2.79	$\sigma_{sy}^2$	0.00
$(gd)_{ik}$ 's	18	5.01	$\sigma_{gd}^2$	-0.03
$(sd)_{ijk}$ 's	70	3.59	$\sigma_{sd}^2$	0.13
$(gdy)_{ikt}$ 's	18	5.74	$\sigma_{gdy}^2$	0.12 <sup>b</sup>
$(sdy)_{ijkt}$ 's	70	2.79	$\sigma_{sdy}^2$	0.09
Error	528	2.53	$\sigma^2$	2.53

<sup>a</sup>Denotes significance at 1 percent level.

<sup>b</sup>Denotes significance at 5 percent level.

and  $\text{Cov}(\text{HS})$  and the estimates of  $\sigma_A^2$ ,  $\sigma_D^2$ , and the interaction of additive and dominance effects with locations and years are presented in Table 19.

The ratio of  $\sigma_D^2/\sigma_A^2$  was 9.51 for the combined analysis as compared to a ratio very close to unity for the average of the six individual

Table 19. Components of genotypic variance for the combined analysis

Component	Estimate <sup>a</sup>	Component	Estimate <sup>a</sup>
Cov(HS)	.03	$\sigma^2_A$	.09
Cov(FS)	.54	$\sigma^2_D$	.84
Cov(HS) <sub>d</sub>	-.03	$\sigma^2_{Ad}$	-.06
Cov(FS) <sub>d</sub>	.06	$\sigma^2_{Dd}$	.19
Cov(HS) <sub>y</sub>	.31	$\sigma^2_{Ay}$	.82
Cov(FS) <sub>y</sub>	.61	$\sigma^2_{Dy}$	.00
Cov(HS) <sub>dy</sub>	.10	$\sigma^2_{Ady}$	.27
Cov(FS) <sub>dy</sub>	.27	$\sigma^2_{Ddy}$	.13

<sup>a</sup>Coded by 3 decimals to the right.

experiments. The additive genetic variance comprised only 9.5 percent of the total genotypic variance in the absence of epistacy.

Although no tests of significance were applied to the estimates of  $\sigma^2_A$ ,  $\sigma^2_D$ , and the interactions of additive effects and dominance deviations with locations and years, an indication of their relative importance was obtained from the tests of significance on  $\sigma^2_g$ ,  $\sigma^2_s$ , and their interactions.

The low value for  $\sigma^2_g$ , which is essentially a measure of the additive genetic variance, indicated that  $\sigma^2_A$  does not deviate

significantly from zero. The significant value of  $\sigma^2_{\text{g}}$  gave evidence that  $\text{Cov}(\text{FS})$  and consequently the value of .84 for  $\sigma^2_{\text{D}}$  was a significant deviation from zero.

The significant value of  $\sigma^2_{\text{gy}}$  indicated a significant deviation from zero for  $\sigma^2_{\text{Ay}}$ , and no evidence of an interaction of dominance deviations with years. The value of .27 for the three factor interaction,  $\sigma^2_{\text{dy}}$ , deviated significantly from zero as supported by the approximate F-test of  $\sigma^2_{\text{gdy}}$ .

If relative comparisons of the variance components were made there was no evidence of an interaction of additive effects with locations, but with years and in the three factor interaction, the interactions with additive effects were larger than those with dominance deviations. The interaction of dominance deviations with locations was the only comparison where dominance deviations showed a greater interaction than additive effects.

## DISCUSSION

Each method which has been proposed for studying covariances among relatives has been of value only within the limits of the assumptions made. Most of the methods proposed have assumed the action of genes to be additive. The theoretical expectations of variances and covariances of relatives were obtained for a single locus and then summed over all loci.

As indicated by Kempthorne (37) for the diallel table where the parents are homozygous lines obtained at random from a random mating population, it appears that ignoring epistacy in the estimation of genotypic components may place a severe restriction on interpretation of the results. Extending the theory to an arbitrary degree of inbreeding in this thesis, for the crosses only, it appears that the presence of epistacy could bias the results considerably.

Kempthorne showed how the estimates obtained from a diallel table could be used under the assumption of no dominance and no epistacy. A test for dominance was possible assuming no epistacy. By extending these results to an arbitrary degree of inbreeding the bias caused by epistacy can be obtained from a properly designed experiment.

Estimates of genotypic components of variance are of interest to a plant breeder because methods of breeding differ depending upon the

types of gene action assumed. Selection for new inbred lines normally begins in an open-pollinated variety, synthetic variety, or some other group of material with a wide genetic base, with the goal to obtain superior combining lines by the most efficient procedure. Basically the estimates of the desired components should be made relative to some population to indicate the type of breeding procedure for maximum progress. The estimates obtained apply only to the particular population involved.

In the procedure used by Hayman (20,21) and Jinks (28,29) a group of arbitrary lines were intercrossed and preliminary examination was made for the presence of non-allelic gene interaction. Any lines which gave an indication of this interaction were removed from the analysis and only the remaining lines were considered in the estimate of dominance. The lines had survived previous testing and represented a highly selected sample. Because of the effects of selection, the results cannot be considered as representing any real population.

In order that estimates of components of genotypic variance be of practical use to a plant breeder the results must apply to some population. After determining the types of gene action operating in the population, breeding methods are used which allow maximum expression of the types of gene action present. This knowledge is combined with the estimates of the means and with measurements of agronomic characters to eventually lead to improved varieties.

The use of the top-cross test for preliminary evaluation of inbred lines of corn assumes that there is a large amount of additive genetic variance present in the original population. Sprague and Tatum (57) showed that estimates of  $\sigma_g^2$  were larger than  $\sigma_s^2$  for a group of unselected lines, indicating that general combining ability was relatively more important. General combining ability was assumed to imply additive gene action with a large value arising when a line is a better or poorer combiner than the lines it is compared with.

Where the lines had been subjected to previous testing, Sprague and Tatum (57) and Rojas and Sprague (50) found that  $\sigma_s^2$  became relatively more important than  $\sigma_g^2$ . Lines remaining from previous selection have had some of the differences in additive effects removed, and dominance and epistatic effects are relatively more important.

In the present experiment there appears to be no consistent trend in the relative magnitudes of  $\sigma_g^2$  and  $\sigma_s^2$ . In five of six experiments, estimates of both  $\sigma_g^2$  and  $\sigma_s^2$  were significant, indicating that for this group of unselected material there is evidence for both additive and nonadditive gene action.

The values of  $\sigma_g^2$  and  $\sigma_s^2$  can be used for obtaining estimates of  $\sigma_A^2$ ,  $\sigma_D^2$ , and  $\sigma_G^2$  under the assumption of no epistacy. There is no previous evidence to indicate the justification of this assumption, and any interpretation of the data is dependent upon it.

The average degree of dominance can not be obtained from the present

data by the method of Comstock and Robinson (7) as  $2\sigma_D^2/\sigma_A^2$ . This ratio measures the average degree of dominance only when the gene frequency of all segregating loci is one-half. In a random mating population the exact gene frequencies are unknown, so the average degree of dominance can not be obtained by this formula.

Hull (25), Brieger (2), and Crow (8) suggested overdominance as the reason intra-variety selection for yield in open-pollinated varieties of corn was ineffective. If the estimate of additive genetic variance obtained from open-pollinated varieties is small relative to the estimate of dominance variance, this would be considered as support of the overdominance hypothesis.

Hull (25) assumed that overdominance effects are of major importance in contrast to partial or complete dominance. Therefore, past selection for yield in a population would have been for a heterozygote, maintaining both alleles in the population in intermediate frequencies. This would cause the majority of the genotypic segregation to arise from dominance deviations. The relatively important values of  $\sigma_A^2$  in relation to the total genotypic variance in the present experiments do not support this hypothesis.

Robinson et al. (48) presented theoretical calculations which indicated that the ratio  $\sigma_D^2/\sigma_A^2$  could vary within wide limits depending upon the degree of dominance and the population frequency of the more favorable allele. Ratios in excess of 1.0 could be obtained with complete dominance,  $a = 1$ . For example, if  $a = 1$ , and the frequency of the more

favorable allele is 0.9, the theoretical ratio of  $\sigma_D^2/\sigma_A^2$  is 4.5. The observed estimate they obtained for yield was .52 in one variety and .33 in another variety.

The ratios of  $\sigma_D^2/\sigma_A^2$  obtained in the present individual experiments were higher than the estimates of Robinson et al., varying from 0.144 to 2.49 in the six experiments. Estimates of the relative magnitude of  $\sigma_A^2$  and  $\sigma_D^2$  support preliminary testing of lines for general combining ability followed by evaluation of specific combinations.

In the combined analysis, however, the ratio of  $\sigma_D^2/\sigma_A^2$  increased to 9.51, a consequence of the reduction of  $\sigma_A^2$  to very near zero. As shown by Robinson, et al., a ratio this large is possible for complete dominance with the gene frequency of the favorable allele near 1.0, or with overdominance at some intermediate gene frequency.

There was no indication of interaction of dominance deviation comparisons with locations or years, while there was considerable interaction of additive gene comparisons with years. The additive genetic variance is defined to be the variance removed by linear regression on the numbers of the possible genes present in the individuals. It appears that there was a shift in the effects of the genes from 1952 to 1953, and consequently a shift in the regression, such that in the combined analysis the average estimate of  $\sigma_A^2$  is almost negligible. The shift in expression of the genes between years is supported by the magnitude of  $\sigma_{Ay}^2$ , which arises from interaction of the additive comparisons of genes with years.



This sizable interaction of  $\sigma^2_{Ay}$  may be an indication of the reason for the failure to obtain large numbers of lines from breeding material which are superior to existing lines in hybrid combinations. Lines which combine poorly in top-cross tests in a given year are discarded and no further evaluation is given to them. If the interaction of additive gene effects with years is large, the only lines which survive testing after several years are those which interact the least with years.

An interaction was not observed for additive effects with locations. However, comparing the variance component for years with that for locations, it appears that the two years provided a more extreme range of environmental conditions than did the three locations. No suggestion is offered for the absence of interaction of dominance deviations with years in contrast to the sizable value for additive effects by years.

The extreme fluctuation in the ratio of the estimates of  $\sigma^2_A$  and  $\sigma^2_D$  between experiments indicate the limitations of these estimates when they are derived from an experiment grown at one location in a single year.

The interaction component estimates are not in close agreement with the results of Rojas and Sprague (50). In their experiments specific effects interacted much more with both locations and years than did general effects.

The major difference in the conduct of the two experiments is in the degree of inbreeding of the parents. Rojas and Sprague considered a diallel table where the parents were completely homozygous. Any

difference between the same entry in different replicates was true replication as all individuals within a plot were expected to be identical genetically. Therefore, in the absence of epistacy,  $\sigma_g^2$  was a direct measure of  $\frac{1}{2}f^2\sigma_A^2$ , and  $\sigma_s^2$  was a measure of  $f^2\sigma_D^2$ .

Each plot in the present experiments represented a heterogeneous population as compared to an inbred plot where every individual is identical genetically. A previous study by Sprague and Federer (56) indicated that the variety x location and variety x year interactions were greater for material which was genetically uniform than for genetically diverse material. It seems possible that in the present experiment the interactions were small because of the heterogeneity of the material within a plot.

It would be useful to determine the magnitude of the interaction of additive gene effects with years over a wider sample of years. If this interaction would remain large in relation to the other components, it would indicate that extensive top-cross testing over several years would be necessary to select a group of lines superior in additive gene effects from that population.

The problem of linkage has not been considered in the present study and it may be important in several aspects of the thesis. Since the original population was a synthetic from 16 lines, the assumption that the source material is a random mating population in equilibrium is probably reasonable unless natural selection is a potent force. After

a number of generations of open-pollination the coupling and repulsion double heterozygotes should be equally frequent unless there is natural selection for a genetic intermediate as proposed by Mather (40), which would lead to an increase of repulsion double heterozygotes over coupling types.

Even if the population is in linkage equilibrium these linkages will affect the covariances as shown by Cockerham (6) for the case when  $F$  equals zero. Assuming no position effects, the covariances between relatives where one is an ancestor of the other are not affected by linkages. However, for full-sibs and half-sibs, linkage causes a bias in the epistatic components of variance, but not in the additive or dominance components.

Estimates of all of the theoretical genotypic variances and covariances could not be obtained from experimental data. It does not appear that there is a good set of diallel cross data in the literature for obtaining all of these estimates. The data used in the present experiment were on plot totals only, limiting the estimates to the variance between sire means and the variance between family means.

The solution to the problem of obtaining estimates of epistatic variances, as well as additive and dominance variances, can be accomplished by making use of different levels of inbreeding,  $F$ . Beginning with a random mating population, a set of parents should be obtained at random, and then selfed without artificial selection for a specified number of generations. After obtaining the desired inbred levels, all possible

intercrosses would be made between the parents within a given level of inbreeding. Individual plant data would enable estimates of variance within families and variance within sires to be obtained in addition to variance between sire means and between family means. The estimates obtained from these different diallel tables would be combined to permit estimation of the components of additive, dominance, and epistatic variances.

Until estimates of these epistatic components are obtained it is impossible to make direct conclusions from any set of data. For example, the estimate of  $\sigma_A^2$  from  $\text{Cov}(\text{HS})$  is an overestimate in the present analysis if the contribution of additive x additive variance is of a sizable magnitude. The presence of any of the additive x additive, additive x dominance, or dominance x dominance components would serve to reduce the observed estimate of  $\sigma_D^2$ . The additive x dominance and dominance x dominance components appear in the variance from which  $\sigma_D^2$  is estimated, but not in the variance from which  $\sigma_A^2$  is estimated. The presence of either of these components of epistatic variance will cause the present estimate of  $\sigma_D^2$  to be an overestimate relative to  $\sigma_A^2$ .

## CONCLUSIONS AND SUMMARY

1. The theoretical covariances between full-sibs and half-sibs are presented for diploid organisms for the general case of arbitrary alleles, arbitrary inbreeding, and arbitrary epistacy. The problem is developed for a single locus with two alleles, extended to a single locus with arbitrary alleles, and finally extended to the general case for two loci with an indication of the results for arbitrary loci.

2. The expectations of variances and covariances of inbred populations are expressed as functions of the original random mating population from which the inbred progeny are obtained at random. The variances obtained are variance between sire means, variance within sires, variance between family means, variance within families, and total genotypic variance.

3. The analysis of variance and expected mean squares are presented for the diallel cross. Theoretical estimates of general and specific combining ability are obtained in terms of additive, dominance, and the two-factor epistatic variances.

4. The variance of general combining ability contains a portion of the additive x additive epistatic variance in addition to additive genetic variance. The variance of specific combining ability contains dominance, additive x additive, additive x dominance, and dominance x dominance variances, independent of additive genetic variance.

5. Estimates of the components of additive and dominance variances and the interactions of additive effects and dominance deviations with locations and years were obtained from the analysis of diallel crosses at three districts for a two year period. Since only plot totals were available, the estimates of variance between sire means and variance between family means were the only variances which could be estimated from the data. The assumption of no epistacy was necessary to allow estimation of the components of additive and dominance variances.

6. In the analysis of individual experiments the relative magnitude of  $\sigma_A^2$  and  $\sigma_D^2$  fluctuated widely between experiments. The averages of these estimates for the six experiments were of approximately equal magnitude, indicating there were both additive and dominance effects in the original population.

7. In the analysis combined over locations and years the estimate of  $\sigma_A^2$  was almost negligible and the estimate of  $\sigma_D^2$  was relatively large. The estimate of the interaction component  $\sigma_{Ay}^2$  was very large relative to the other interaction components, a result of a shift in the effects of the genes from 1952 to 1953. This shift in expression of the genes accounted for the small estimate of  $\sigma_A^2$  in the combined analysis, since the additive genetic effects averaged to almost zero. This would have considerable relevance to breeding and testing procedures if it is substantially correct.

8. A brief discussion is given to the complications of linkage.

Considerations are directed towards populations which are in linkage equilibrium and those which are not.

9. Before adequate interpretation can be given to studies on quantitative inheritance, information is needed on the importance of epistacy. A method is suggested which utilizes the variances derived in this thesis to obtain estimates of epistatic components of genotypic variance.

## LITERATURE CITED

1. Anderson, V. L. and Kempthorne, O. A model for the study of quantitative inheritance. *Genetics* 39:883-898. 1954.
2. Brieger, F. G. The genetic basis of heterosis in maize. *Genetics* 35:420-445. 1950.
3. Bruce, A. B. The Mendelian theory of heredity and the augmentation of vigor. *Science* 32:627-628. 1910.
4. Byrd, W. P. Genetic and environmental variances in segregating and non-segregating maize populations. Unpublished Ph.D. Thesis. Ames, Iowa, Iowa State College Library. 1955.
5. Cockerham, C. C. An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. *Genetics* 39:859-882. 1954.
6. \_\_\_\_\_. Effects of linkage on the covariances between relatives. *Genetics* 41:138-141. 1956.
7. Comstock, R. E. and Robinson, H. F. The components of genetic variance in populations of biparental progenies and their use in estimating the average degree of dominance. *Biometrics* 4:254-266. 1948.
8. Crow, J. F. Alternate hypothesis of hybrid vigor. *Genetics* 33:477-487. 1948.
9. Davenport, C. B. Degeneration, albinism, and inbreeding. *Science* 28:454-455. 1908.
10. Dickinson, A. G. and Jinks, J. L. A generalised analysis of diallel crosses. *Genetics* 41:65-78. 1956.
11. East, E. M. A Mendelian interpretation of variation that is apparently continuous. *Am. Nat.* 44:65-82. 1910.
12. \_\_\_\_\_. Heterosis. *Genetics* 21:375-397. 1936.
13. \_\_\_\_\_, and Hayes, H. K. Heterozygosis in evolution and in plant breeding. U. S. Dept. Agr., Bur. Pl. Ind. Bul. 243:1-58. 1912.



14. Federer, W. T. and Sprague, G. F. A comparison of variance components in corn yield trials: I. Error, tester x line, and line components in top-cross experiments. *Jour. Am. Soc. Agron.* 39:453-463. 1947.
15. Fisher, R. A. The correlation between relatives on the supposition of Mendelian inheritance. *Trans. Roy. Soc. of Edinburgh* 52:399-433. 1918.
16. Gardner, C. O., Harvey, P. H., Comstock, R. E., and Robinson, H. F. Dominance of genes controlling quantitative characters in maize. *Agron. Jour.* 45:186-191. 1953.
17. Green, J. M. Relative value of two testers for estimating top cross performance in segregating maize progenies. *Jour. Amer. Soc. Agron.* 40:45-57. 1948.
18. Griffing, B. Analysis of quantitative gene action by constant parent regression and related techniques. *Genetics* 35:303-321. 1950.
19. Hardy, G. H. Mendelian proportions in a mixed population. *Science* 28:49-50. 1908.
20. Hayman, B. I. The analysis of variance of diallel tables. *Biometrics* 10:235-244. 1954.
21. \_\_\_\_\_. The theory and analysis of diallel crosses. *Genetics* 39:789-809. 1954.
22. \_\_\_\_\_ and Mather, K. The description of genic interaction in continuous variation. *Biometrics* 11:69-82. 1955.
23. Horner, T. W. Non-allelic gene interaction and the interpretation of quantitative genetic data. Unpublished Ph.D. Thesis. Raleigh, North Carolina, North Carolina State College Library. 1952.
24. \_\_\_\_\_ and Kempthorne, O. The components of variance and the correlations between relatives in symmetrical random mating populations. *Genetics* 40:310-320. 1955.
25. Hull, F. H. Recurrent selection for combining ability in corn. *Jour. Amer. Soc. Agron.* 37:134-145. 1945.
26. \_\_\_\_\_. Recurrent selection and overdominance. In Gowen, J. W., ed. *Heterosis*. p. 451-473. Ames, Iowa, Iowa State College Press. 1952.
27. Jennings, H. S. The numerical results of diverse systems of breeding, with respect to two pairs of characters, linked or independent, with special relation to the effects of linkage. *Genetics* 2:97-194. 1917.

28. Jinks, J. L. The analysis of continuous variation in a diallel cross of Nicotiana rustica varieties. *Genetics* 39:767-788. 1954.
29. \_\_\_\_\_. A survey of the genetical basis of heterosis in a variety of diallel crosses. *Heredity* 9:223-238. 1955.
30. Jones, D. F. Dominance of linked factors as a means of accounting for heterosis. *Genetics* 2:466-479. 1917.
31. Keeble, F. and Pellow, G. The mode of inheritance of stature and of time of flowering in peas (Pisum sativum). *Jour. Genetics* 1:47-56. 1910.
32. Keller, K. R. A comparison involving the number of, and relationship between, testers in evaluating inbred lines of maize. *Agron. Jour.* 41:323-331. 1949.
33. Kempthorne, O. The design and analysis of experiments. N. Y., John Wiley and Sons, Inc. 1952.
34. \_\_\_\_\_. The correlation between relatives in a random mating population. *Proc. Roy. Soc. B* 143:103-113. 1954.
35. \_\_\_\_\_. The theoretical values of correlations between relatives in random mating populations. *Genetics* 40:153-167. 1955.
36. \_\_\_\_\_. The correlations between relatives in inbred populations. *Genetics* 40:681-691. 1955.
37. \_\_\_\_\_. The theory of the diallel cross. (To be published in *Genetics*, July, 1956).
38. \_\_\_\_\_. An introduction to genetic statistics. (Unpublished manuscript of book in preparation.) 1956.
39. Malécot, G. *Les mathématiques de l'hérédité*. Paris, Masson et Cie. 1948.
40. Mather, K. *Biometrical genetics*. N. Y., Dover Publ., Inc. 1949.
41. \_\_\_\_\_ and Vines, A. The inheritance of height and flowering time in a cross of Nicotiana rustica. In Reeve, E. C. R. and Waddington, C. H., eds. *Quantitative inheritance*. p. 49-79. London, Her Majesty's Stationery Office. 1952.
42. Matzinger, D. F. Comparison of three types of testers for the evaluation of inbred lines of corn. *Agron. Jour.* 45:493-495. 1953.
43. Nilsson-Ehle, H. *Kreuzungsuntersuchungen an Hafer und Weizen*. Lund Univ. Arsskr. N. F. Afd. 2. Bd. 5:1-122. 1909.

44. Pearson, K. On a generalized theory of alternative inheritance with special reference to Mendel's laws. *Phil. Trans. Roy. Soc., A.* 203:53-86. 1904.
45. Powers, L. Determining scales and the use of transformations in studies on weight per locule of tomato fruit. *Biometrics* 6:145-163. 1950.
46. \_\_\_\_\_. Gene analysis by the partitioning method when interactions of genes are involved. *Bot. Gaz.* 113:1-23. 1951.
47. Robinson, H. F., Comstock, R. E., and Harvey, P. H. Estimates of heritability and the degree of dominance in corn. *Agron. Jour.* 41:353-359. 1949.
48. \_\_\_\_\_, \_\_\_\_\_, and \_\_\_\_\_. Genetic variances in open pollinated varieties of corn. *Genetics* 40:45-60. 1955.
49. Rojas, B. A. Analysis of a group of experiments on combining ability in corn. Unpublished M.S. Thesis. Ames, Iowa, Iowa State College Library. 1951.
50. \_\_\_\_\_ and Sprague, G. F. A comparison of variance components in corn yield trials: III. General and specific combining ability and their interaction with locations and years. *Agron. Jour.* 44:462-466. 1952.
51. Satterthwaite, F. E. An approximate distribution of estimates of variance components. *Biometrics Bul.* 2:110-114. 1946.
52. Shull, G. H. The composition of a field of maize. *Amer. Breeders Assoc.* 4:296-301. 1908.
53. \_\_\_\_\_. The genotypes of maize. *Amer. Nat.* 45:234-252. 1911.
54. \_\_\_\_\_. Duplicate genes for capsule form in *Bursa bursa-pastoris*. *Zeitschr. für Induk. Abstamm. u. Vererb.* 12:97-149. 1914.
55. Sprague, G. F. Heterosis. In Loomis, W. E., ed. *Growth and differentiation in plants.* p. 113-136. Ames, Iowa, Iowa State College Press. 1953.
56. \_\_\_\_\_ and Federer, W. T. A comparison of variance components in corn yield trials: II. Error, year x variety, location by variety, and variety components. *Agron. Jour.* 43:535-541. 1951.
57. \_\_\_\_\_ and Tatum, L. A. General vs. specific combining ability in single crosses of corn. *Jour. Amer. Soc. Agron.* 34:923-932. 1942.

58. Weinberg, W. Über Vererbungsgesetze beim Menschen. Zeit. Ind. Abst. Ver. 1:377-392, 440-460; 2:276-330. 1908.
59. \_\_\_\_\_. Weitere Beiträge zur Theorie der Vererbung. Arch. Rass. u. Ges. Biol. 7:35-49, 169-173. 1910.
60. Wentworth, E. N. and Remick, B. L. Some breeding properties of the generalized Mendelian population. Genetics 1:608-616. 1916.
61. Wright, S. Systems of mating, I - V. Genetics 6:111-178. 1921.
62. \_\_\_\_\_. The analysis of variance and the correlation between relatives with respect to deviations from optimum. Jour. Genetics 30:243-256. 1935.
63. \_\_\_\_\_. The genetics of quantitative variability. In Reeve, E. C. R. and Waddington, C. H., eds. Quantitative inheritance. p. 5-41. London, Her Majesty's Stationery Office. 1952.
64. Yates, F. Analysis of data from all possible reciprocal crosses between a set of parental lines. Heredity 1:287-301. 1947.
65. Yule, G. U. On the theory of inheritance of quantitative compound characters on the basis of Mendel's laws--a preliminary note. Report of Third International Conference on Genetics. p. 140-142. 1906.

## ACKNOWLEDGEMENT

The writer is most grateful to Dr. G. F. Sprague for guidance throughout the course of the investigation. Indebtedness is expressed to Professor O. Kempthorne for his assistance in the theoretical genetical derivations and in the statistical analysis and for use of his unpublished manuscripts. Appreciation is also expressed to both of these men for helpful suggestions in the preparation of the manuscript.